Single Molecule Force Spectroscopy with Biological Tools

Wolfgang Bernhard Ott

München 2017

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Single Molecule Force Spectroscopy with Biological Tools

Wolfgang Bernhard Ott

Dissertation
an der
Ludwig-Maximilians-Universität München
Fakultät für Physik

vorgelegt von
Wolfgang Bernhard Ott
aus
Eschenbach in der Oberpfalz

München, den 15. November 2017

Erstgutachter: Prof. Dr. Hermann E. Gaub
Zweitgutachterin: Prof. Dr. Claudia Veigel
Tag der mündlichen Prüfung: 20.12.2017

Zusammenfassung

Im Fokus dieser Arbeit steht die Mechanobiologie der extrazellulären Maschinerie des Cellulosoms. Cellulosome sind Multiprotein-Komplexe und werden von Pilzen und Bakterien sekretiert um enzymatisch Hemicellulose, Lignin und Cellulose effizient zu hydrolysieren. Das Cellulosom ist ein vielversprechendes Modellsystem um verschiedene mechanoresistente Rezeptor:Liganden-Interaktionen zu untersuchen, aber auch um die Dynamik und Komplexität eines Proteinnetzwerks zu verstehen.
Einzelmolekülkraftspektroskopie (SMFS) ermöglicht es individuelle, cellulosomale Cohesin:Dockerin Interaktionen zu betrachten und so Rückschlüsse auf das ganze Cellulosom zu ziehen. Dabei werden einzelne Proteine oder Rezeptor:Liganden-Interaktionen mit einem Rasterkraftmikroskop (AFM) mechanisch belastet und deren Entfaltungskraft aufgezeichnet, indem diese zwischen einer Glasoberfläche und der Messnadel des AFMs fixiert werden.
In dieser Arbeit wurde die mechanische Stabilität des Cohesin:Dockerin Typ III Komplexes aus *Ruminoccous flavefaciens* untersucht. Dabei wurden Abrisskräfte von ca. 650 pN beobachtet, womit diese zu den höchsten gemessenen Abrisskräften eines einzelnen Rezeptor:Liganden-Systems gehören.
Diese hohen Kräfte führen zur Überdehnung von Polyethylenglykol (PEG), Polymere die bei AFM-Experimenten standardmäßig als Abstandhalter eingesetzt werden. Die Konformation von PEG wechselt, bei einer Belastung von über 350 pN, von einer gestauchten trans-trans-gauche Konformation in eine komplette trans Konformation, welche die Konturlänge verändert.
Um sich dieser Angelegenheit anzunehmen wurde in dieser Arbeit mit biologischen Polymeren gearbeitet, sog. Elastin-ähnliche Polypeptide (ELPs). ELPs können mit gängigen posttranslationalen Proteinligationsmethoden mit dem Zielprotein ortsspezifisch verknüpft werden und so in AFM-Versuchen als Abstandhalter eingesetzt werden. Die Peptidbindung der ELPs beeinflußt die eigentliche Dehnung der Zielproteine nicht mehr. Außerdem können ELPs auf Grund ihrer biologischen Natur leichter angepasst werden. So ist es möglich *up- und downstream* des ELP-Gens Sequenzen anzubringen, welche unterschiedliche, ortsspezifische Oberflächenanbindungen ermöglichen.
Im Weiteren Verlauf der Arbeit wurde der Durchsatz bei SMFS-Experimenten verbessert. Mit Hilfe zellfreier Expression konnten im kleinen Maßstab mehrere verschiedene Proteine synthetisiert, räumlich getrennt auf einer Glasoberfläche immobilisiert und im Anschluss mit der gleichen Messnadel vermessen werden. Somit können die absoluten Entfaltungskräfte direkt miteinander verglichen werden.
Im letzten Teil der Arbeit wurden spezifische Proteinfunktionalisierungsstrategien für SMFS- und Bulk-Affinitätsexperimente etabliert. So ist es nun möglich die gleiche Anbinungsstrategie in oberflächengebundenen Affinitätsassays als auch in Einzelmolekülkraftspektroskopiemessungen anzuwenden.
Zusammenfassend beschreibt diese Arbeit die Etablierung und Optimierung biochemischer Werkzeuge, zur ortspezifischen und funktionellen Anbindung von Proteinen des Cellulosoms in der SMFS und in biophysikalischen Bulkexperimenten, um wertvolle Einsichten über deren Bindemechanik und Stabilität zu gewinnen.

Abstract

The mechanobiology of the cellulosome, an extracellular multiprotein machinery was the focus of this thesis. Cellulosomes are secreted by fungi and bacteria to efficiently hydrolyze hemicellulose, lignin and cellulose, using different enzymes. The cellulosome is a promising model system to investigate mechanoresistant receptor:ligand interactions as well as to understand the dynamics of a complex protein network.
Single molecule force spectroscopy (SMFS) allows the examination of individual cellulosomal cohesin:dockerin interactions providing insights about the whole cellulosome. By immobilizing single proteins or receptor:ligand interactions between an atomic force microscope (AFM) cantilever and a glass surface, they can be stretched mechanically with an AFM to record unfolding behavior.
In this thesis, the mechanical stability of the type III cohesin:dockerin complex of *Ruminoccous flavefaciens* was probed. Rupture forces of ca. 650 pN could be observed, which rank amongst the highest rupture forces of a single receptor:ligand interaction. Reaching these high forces leads to an overstretching of the standard linker polymer polyethylene glycol (PEG) used in AFM SMFS. In an AFM experiment exceeding 350 pN, PEG undergoes a transition from the compressed trans-trans-gauche conformation to the all-trans conformation and thereby changes the contour length.
To address this issue biological polymers, so called Elastin-like polypeptides (ELPs) were explored. ELPs are linked to a protein of interest with established post-translational ligation methods and can be employed as linkers for SMFS. The ELP peptide bonds no longer distort the stretching of the protein of interest. The biological nature of ELPs allows to easily customize them by incorporating sequences for post-translational modification up- and/or downstream of the gene.
Furthermore, this thesis improved the throughput and comparability of AFM experiments. Using cell-free expression several different proteins were produced in a small scale and subsequently immobilized spatially separated on a glass slide, enabling their measurement with a single cantilever. The resulting unfolding events were comparable in absolute forces.
In the final part of this thesis, strategies for site-specific protein immobilization for SMFS and bulk-affinity experiments were established. It is now possible to apply the same surface immobilization strategies in surface-bound, label-free affinity assays and in SMFS experiments.
In summary, this thesis establishes and optimizes tools from biochemistry for site-specific and functional immobilization of different cellulosomal proteins in SMFS and biophysical bulk assays, to gain insight in their binding mechanisms and mechanostability.

Contents

Chapter 1

Introduction

A hallmark of life is its capability to respond to changing environments. Cells, the building blocks of organisms adapt with microscopic changes, as the environment changes on a macroscopic level. These changes happen dynamically, orchestrated by interactions between biomolecules: DNA, lipids and proteins. Their interplay senses, processes and responds to a vast variety of environmental changes.
Each biomolecule fulfills specific tasks so that the cell as a whole is able to respond efficiently. DNA delivers the basic blueprints for the molecular scheme [1], [2]. Lipids create reaction compartments to facilitate concentration gradients thus fueling processes and enable spatially separated reactions [3]. Proteins are the most versatile biomolecules as they provide a whole portfolio of tasks: For example, they bind, transport and release compounds or open and close channels for agents to diffuse *via* Brownian motion [4]. In form of the cytoskeleton they provide stability to a cell, or transfer and buffer force in muscle cells [5], [6]. As enzymes, they catalyze crucial chemical reactions by lowering the energy barrier for compound transformation.
In order to understand these complex protein networks, it is instrumental to isolate individual interaction partners for analysis. Independent of their function, proteins realize their task by close range, mostly non-covalent interactions [7]. Bottom-up approaches have proven to be useful to examine them. Single interactions are studied, and once understood, assembled into a larger picture like a jigsaw puzzle - piece by piece.
Scanning probe atomic force microscopy (AFM) was shown to be a viable tool to visualize processes on a nanometer scale. Besides AFM imaging, single molecule force spectroscopy (SMFS) has evolved as a technique to probe the mechanostability of proteins and receptor:ligand interactions. SMFS reveals the stability or interaction of proteins in a real quantitative manner, measuring forces down to a low pN-range [8].
In order to obtain accurate results, a bottom-up approach should mimic physiological conditions in a simplified setup. Recent developments in bioconjugation allowed to functionalize proteins in a site-directed manner for SMFS experiments to ensure probing of their endogenous binding geometry [9], [10]. Specifically, genetically encoded recognition sequences can be used to covalently fuse proteins to small molecules or other peptide sequences. Advancements in cloning techniques allow the custom construction of DNA sequences encoding the protein of interest with desired modifications [11].
To efficiently measure the biomechanics of a protein of interest it is fused to a binding handle of a receptor:ligand interaction with known unbinding forces. This fusion protein is site-specifically immobilized on a surface. The complementary domain is site-specifically immobilized on the cantilever of the experimental SMFS setup. Upon approaching the cantilever, the complex of receptor and ligand can form. Then, the cantilever is retracted and receptor:ligand complex as well as the protein of interest

are loaded with force. Ideally, the protein domain of interest unfolds first and then receptor and ligand dissociate. This strategy allows to probe several hundreds copies of the same protein of interest sequentially.
A similar approach can be used to study receptor:ligand interactions. In this case the binding handles are exchanged for an unknown receptor:ligand system and are fused to protein domains with known unfolding characteristics (unfolding force distribution, possible sub-steps, added contour lengths), so called fingerprint domains. Fingerprint domains provide unambiguous identification of specific single molecule interactions. Only unfolding traces showing all expected unfolding events of fingerprint domains (*e.g.*, surface and cantilever fusion protein) and complex dissociation are used in the later analysis.
To provide enough conclusive statistics, sufficient characteristic unfolding traces need to be sampled. A robust experimental setup is necessary, refolding fingerprint domains and receptor:ligand interactions with high dissociation forces. High dissociation forces ensure that the complex only ruptures, after the protein of interest unfolded in the experiment [12].
The cellulosome has shown to be a gold mine for the search for robust and high affinity binding handles as well as a model system for protein function and dynamics [13], [14]. The cellulosome is a large multiprotein machinery, secreted by anaerobic bacteria and fungi to hydrolyze the world's most abundant biomass cellulose to glucose [15]. Initially, organisms need to hydrolyze hemicellulose and lignin, to access the high energy polymer cellulose. Lignin is a very complex and heterogeneous compound, which makes it necessary to employ several different enzymes, each specialized for one type of chemical bond. Cellulolytic organisms organize these enzymes by anchoring the extracellular cellulosome to their cell wall. The cellulosome is based on framework proteins (scaffoldins), consisting of several binding domains called cohesins. The scaffoldin often contains a carbohydrate binding domain (CBM), which links cells in close proximity to cellulose. The cohesins on the scaffoldin are recognized by dockerins, a binding domain of most cellulolytic enzymes. Hence, different enzymes are placed in close distance to each other and the substrate to generate a synergistic effect, increasing cellulose hydrolysis.
Consequently cohesin:dockerin interactions are optimized for high stability and specific binding, motivating their investigation by SMFS.
Interdisciplinary approaches between biology, chemistry and physics dissect the cellulosomal machinery, providing insights into the mechanism and developing new synthetic applications, *i.e.* usage of cohesin:dockerin as robust binding handles in SMFS.

Chapter 2

Scientific Context

2.1 How Proteins Generate Mechanical Stability

All known life forms use the same set of 20+2 crucial building blocks - the natural amino acids for protein synthesis. Plus two stands for selenocysteine and pyrrolysine, which are only present in certain organisms [16], [17]. DNA encodes the sequence of amino acids, which then determines protein structure and function. Three subsequent base pairs of DNA (codons) encode one amino acid. Each codon is unique for one amino acid, but because there are more than 20 permutations possible the genetic code is degenerate, *i.e.* one amino acid can be encoded by several different codons [18]. DNA is transcribed into mRNA which in turn is translated by ribosomes into the amino acid sequence (primary structure). The sequence of amino acids determines the folding of the protein [19]. Mostly two main secondary structures exist in protein folding topology [20]: Helical shapes, termed α-helix or sheet like structures, termed β-sheets. α-Helices and β-sheets are connected by variable loop regions which altogether form the final fold of a monomeric protein (tertiary structure). Proteins can also contain more than one subunit, the so called quaternary structure can be homologous (repeat of a same domain) or heterologous (different folded protein domains) [21].
Why do protein domains stay folded, how do multimeric protein domains stabilize each other and how do different proteins fulfill different functions?
Figure 1 gives an overview of the most important interactions within a protein and between a protein and its environment. Amino acids are linked together *via* a condensation reaction and form a covalent peptide bond. In this reaction the α-carboxyl-group of one amino acid reacts with the α-amino group of another amino acid, releasing a water molecule. The resulting linear polymer contains a free amino-group at one end (N-terminus) and a carboxyl-group at the opposite end (C-terminus) [22]. Each amino acid has a unique side chain defining its property. They can be grouped by the characteristics of the side chain: There are six polar/uncharged, six non-polar/aliphatic, three positively charged, two negatively charged and three non-polar/aromatic amino acids.
During initial protein folding the non-polar/hydrophobic amino acids are buried in the protein core (hydrophobic collapse) whereas the polar/hydrophilic amino acids are solvent exposed [23]. During the collapse α-helices and β-sheets form, because neighboring side chains dictate their folding. After the collapse local rearrangements take place and optimize solvent exposure of the globular protein domain, by moving side chains inside (hydrophobic amino acids) or outside (hydrophilic amino acids) of the protein core. Hydrophilic amino acids usually form hydrogen bonds with water molecules at the outer shell of a protein.
Hydrogen bonds play a crucial role in stabilization of α-helices and β-sheets [24]. The oxygen- and nitrogen-atoms of two different amide-groups in the protein backbone

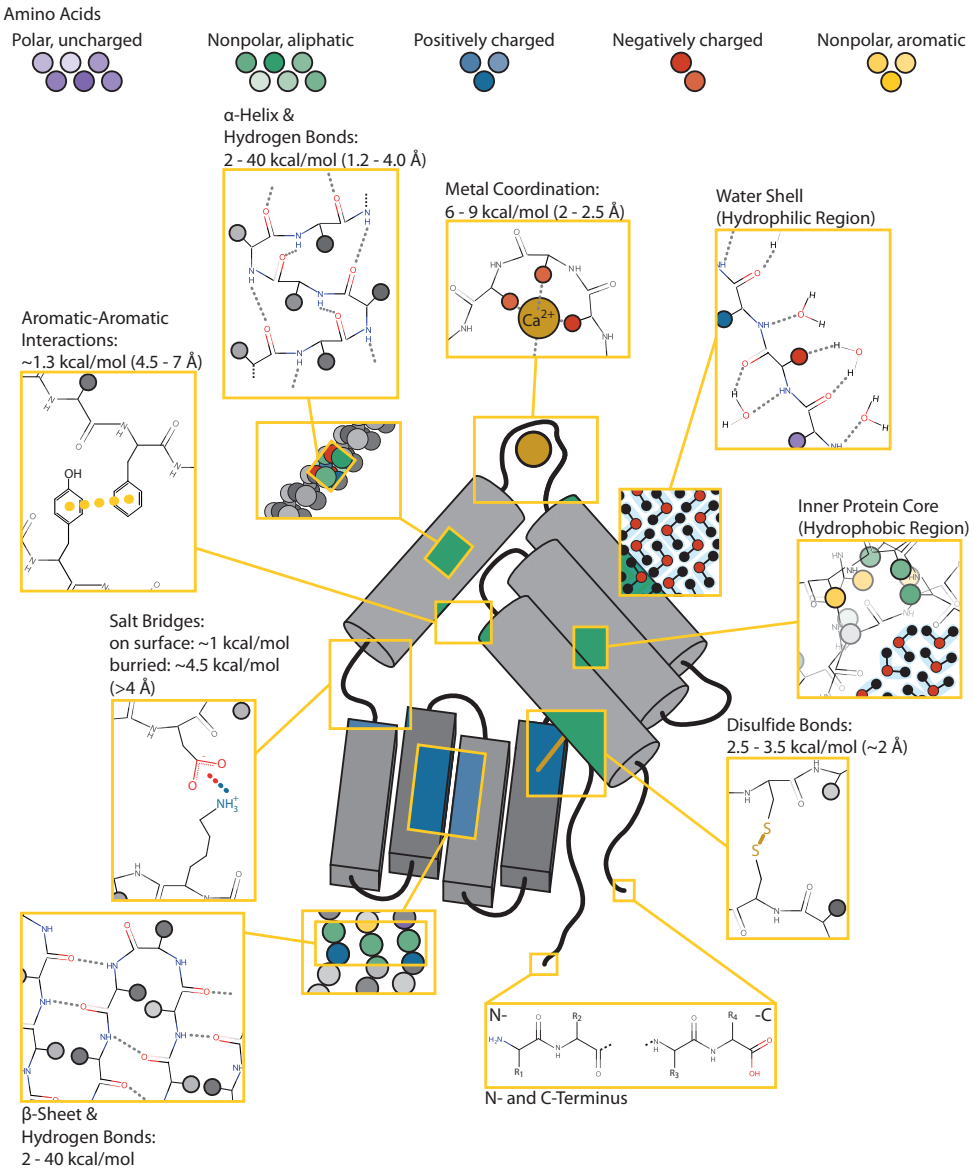


Figure 1. Overview of intra- and inter-molecular interaction of proteins. Schematic and chemical structures of the molecular mechanisms of different interaction types in proteins with corresponding binding distances and energies.

interact with each other, if they are no further than 2.5 Å apart. Different side chains can also contribute to hydrogen bonding, *i.e.* glutamine or asparagine have a primary amine-group which is able to interact with an oxygen-atom.
Negatively charged side chains (*i.e* glutamate or aspartate) can form salt bridges with positively charged side chains of amino acids (*i.e.* lysines, arginines) to stabilize a protein [25]. Salt bridges consist of two types of interactions: An electrostatic interaction between oppositely charged groups and hydrogen bonding between oxygen (carboxyl-group)- and hydrogen (amine-group)-atoms [26], [27]. Salt bridges are mostly formed in the inner core of proteins, since physiological conditions often contain counter-polarized salt ions, which shield the electrostatic charge of solvent-accessible side chains [25], [28], [29].
The aromatic side chains (tryptophan, phenylalanine, or tyrosine) can interact with each other [30], [31]. So called aromatic-aromatic interactions are based on interactions between π-electron rings of two close aromatic side chains [32]. The π- electrons can

also interact with charged residues or CH-groups. However, this is more important for substrate recognition than for protein stabilization [33].
Anorganic matter coordination (cations, anions, and phosphate) is stabilized by charged amino acid side chains, *i.e.* glutamate and aspartate, but also histidines, cysteines and tyrosine play a crucial role [34]. Flexible loops are stiffened by coordination of metal ions, which contributes to the overall rigidity of proteins [35]–[38].
Cysteine is the only natural amino acid which is able to form a covalent bond *via* its thiol side chain. Two cysteines can form a disulfide bond, called cystine, which grants high mechanical stability in proteins by fixing the fold in certain positions [39]–[42].
These manifold interactions contribute to the unique 3D-conformation, but also allow proteins to perform specialized tasks. For example, calmodulin changes its conformation upon calcium binding and triggers a signal cascade [43]. Avidin, a tetrameric protein, complexes a small molecule (biotin) with extremely high affinity in the femtomolar range [44]. Hemoglobin coordinates, transports and releases oxygen [45]. The cytoskeletal proteins actin, tubulin and intermediate filaments provide stability in eukaryotic cells [46]. DNA polymerases recognize double stranded DNA, repair DNA-damage and synthesize DNA [47]. GFP (green fluorescent protein) emits green light upon excitation based on its unique fold, leading to chromophore formation with three close interacting side chains [48].
A model system for the interplay of protein interactions, function and enzymatic activity is the cellulosome, a fungal or bacerial multiprotein complex.

2.2 The Cellulosome - A Model System for Biomechanics and Protein Network Dynamics

Cellulosomes are highly structured multiprotein complexes secreted by anaerobic fungi and bacteria to digest cellulose, hemicellulose and lignin. Cellulose is the most abundant biomaterial on earth [49]. It can be found in all plants, some algae and even some bacteria secrete cellulose fibers.
Cellulose is a linear polymer consisting of only β-1-4-glycosidic linked glucose molecules (**Figure 2**) [50]. Repeats of cellobiose, two linked glucose molecules are subunits of cellulose. Usually cellulose is thousands of cellobiose repeats long [51]. The two ends of each cellulose chain are termed reducing and non reducing end. At the reducing end, the glucose ring can adopt an open conformation and an aldehyde-group is freed, which has reducing potential. Some cellulolytic enzymes are specialized in hydrolyzing reducing or non reducing ends of cellulose.

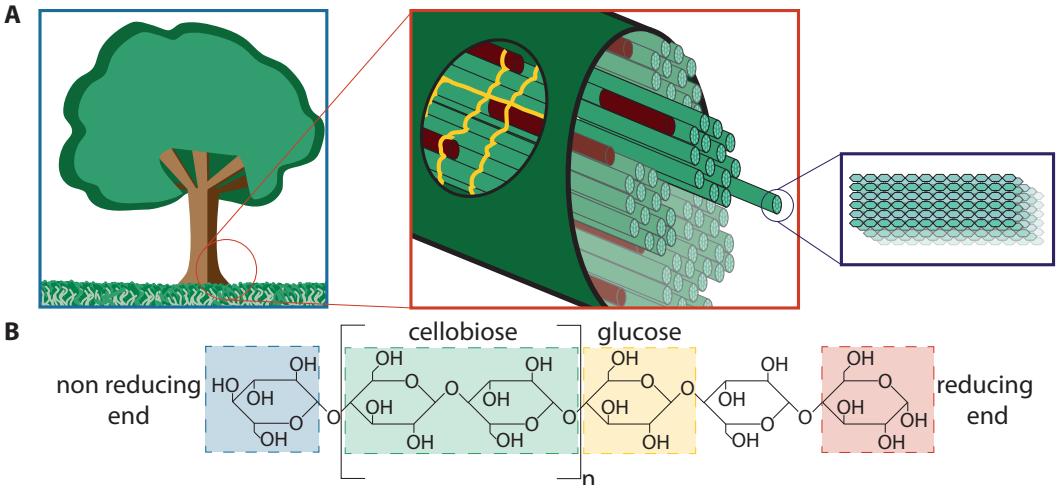


Figure 2. Scheme of the complex nature of cellulose - the most abundant biomaterial on earth. A All plants contain cellulose. The red box shows the macroscopic structure of cellulose (green). The linear polymer is surrounded by hemicellulose (red) and gaps in-between are connected by lignin (yellow). The blue box shows the composition of cellulose a homogeneous, linear polymer. Cellulose microfibrils lay ontop of each other and form a perfectly aligned cellulose fiber. **B** Chemical structure of the cellulose polymer. Cellulose is a polymer consisting of glucose molecules connected *via* a β-1-4-glycosidic bond. The polymer has a non reducing end (blue) and a reducing end (red). The reducing end is able to open the glucose ring and form an aldehyde group, which can act as a reducing agent. Cellobiose is labeled in green, which is the smallest product obtained through cellulosomes. Glucose is highlighted in yellow.

Cellulose is stabilized by intra- and intermolecular hydrogen bonds [52]. Cellulose polymer strands align next to another to form so called microfibrils [51]. Several microfibrils form cellulose fibrils, which again form the macroscopically visible cellulose fibers. If this overall alignment of cellulose fibers is present the cellulose is called crystalline. Whenever the alignment is disturbed, *i.e.* microfibrils are twisted and misaligned they cause disordered parts in the cellulose architecture, termed amorphous regions.
Cellulose is a high-energy polymer, providing an efficient food source for living organisms. However, cellulose is usually shielded from degradation by a protective envelope consisting of hemicellulose and lignin [53]. Hence, to access cellulose different chemical bonds need to be degraded beforehand. Especially lignin is a

heterogeneous polymer, consisting of substances like xylose, mannose, or lichenin [54].
As enzymes are usually highly specific for one chemical reaction a broad portfolio of enzymes is necessary [55]. The discovery of the cellulosome explained how organisms deal with the enormous substrate heterogeneity [56]. The architecture of the extracellular, multienzyme complex is based on a crucial receptor:ligand interaction called cohesin:dockerin (**Figure 3**) [15]. The basic principle of a cellulosome machinery is to link as many enzymes as possible in close proximity to each other. This is realized by framework proteins called scaffoldins. A scaffoldin is a repetitive protein with several cohesin domains and sometimes a CBM. All endogenous cellulolytic enzymes consist of a catalytic domain and a cohesin binding domain called dockerin. The interaction between the enzymatic dockerin and the cohesin on the scaffoldin is classified as type I [57].

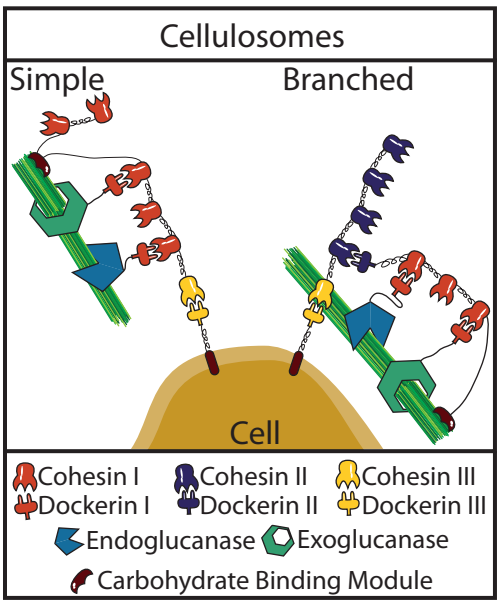


Figure 3. Schematic of cellulosomes. Cell-connected cellulosomes can be divided into two major groups. Simple cellulosomes (left) mostly consist of one big scaffold where enzymes can dock onto, connecting enzymes and cells very close to the substrate. Branched cellulosomes (right) interconnect scaffolds to create a larger surface for enzymes to dock onto.

Two more types of cohesin:dockerin interactions exist in the cellulosome, classified by their position and function. Usually type II and III cohesin:dockerin interactions facilitate the attachment of scaffoldins to cell walls and branching of the cellulosome [58], [59], respectively. Gram-positive bacteria secret a cohesin possessing a LPXTG-motif, which is recognized and covalently linked to the cell wall by a Sortase [60]. The cell wall-anchored cohesin is recognized by a dockerin containing scaffoldin and constitutes the basis for the cellulosomal assembly. Depending on the architecture of the scaffoldin, enzymes can dock or additional branching scaffoldins can bind. The branching process is promoted by specific type II cohesin:dockerin interactions.
Cellulosomes usually consist of scaffoldins with different types of CBMs depending on the substrate and its shape, *i.e* amorphous or crystalline cellulose [61]–[63].

The combination of domains that anchor the cellulosome both to the substrate and the cell enable the localization of enzymes in close proximity to their substrate and allow an efficient uptake of the product by the host cell.
In order to transfer the highly specialized cellulosomal assembly principle to different applications or to understand the underlying mechanisms of the cohesin:dockerin interactions it is necessary to investigate their binding characteristics. Their affinities, determined by SPR (surface plasmon resonance) [64], or ITC (isothermal titration calorimetry) [65] are in the pico- or nanomolar range. Most of the known cohesins and dockerins from different species do not bind each other, hence are highly species-specific. However, due to their high grade of homology some cohesin:dockerin cross-reactivity between different organisms is observed [64].
What defines the specificity of an interaction? Why do two binding partners build strong complexes but do not interact with a third despite their high structural homology?
The cellulosome is an ideal model system to investigate these questions. Single molecule force spectroscopy (SMFS) with nanometer resolution is able to identify the crucial amino acids involved in binding mechanics, generating binding affinity, specificity and their mechanobiology.

2.3 Probing Biomechanics of Single Molecules

Atomic force microscopy (AFM) has been used to study protein mechanics of cellulosomal components. Results showed that binding affinity does not necessarily correlate with mechanical stability. The type III cohesin:dockerin interaction, which exhibits rupture forces around 650 pN [13] is much more mechanostable than the type I cohesin:dockerin interaction with rupture force around 120 pN [66]. In contrast to their binding affinities: Type I (pM range [64]) binds with higher affinity than type III (about 20 nM [59]).
AFM SMFS can link force exerted on a molecule to its function and can supply insights into the binding or folding characteristics on a nanometer scale. AFM experiments investigated the mechanobiology of a scaffoldin from *Clostridium thermocellum*. Valbuena *et al.* could find correlations between stability, function and position. They showed that the mechanical stability of the cohesins was dependent on their position in the scaffold [67]. Depending on the position of the cohesin domain - between cell and CBM ("bridging cohesin") or free after the CBM adhering to the cellulose ("hanging cohesin") - their unfolding force differed. Bridging cohesins unfolded at 400 to 600 pN, hanging cohesins already at forces around 280 pN.
In order to obtain data that can be analyzed, single molecule interactions need to be probed. Typically, a commercially available cantilever is used as a force probe and functionalized with proteins. **Figure 4** shows scanning electron microscope (SEM) images of an Olympus BioLever mini, a cantilever that is commercially available and commonly used in AFM.

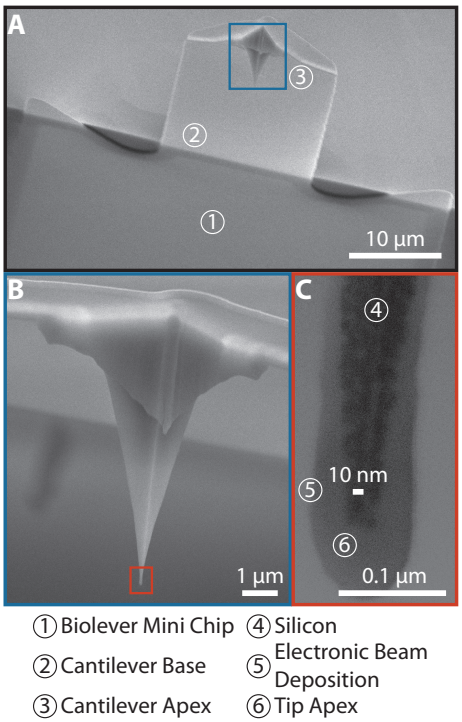


Figure 4. Scanning electron microscope (SEM) pictures of a BioLever Mini. A The structure of a BioLever mini acquired by SEM. The chip (1), the cantilever (2) and the tip of the canilever (3) are depicted. **B** Magnification of the cantilever tip (3). **C** Magnification of the tip apex. The actual apex of the very tip (6) is visible (dark grey), which is roughly 20 nm in size (4). The lighter grey vincinity of the tip is the electronic beam deposition (5), an image artifact generated by the sample preparation procedure for the cantilever imaging. Pictures were kindly provided by Dr. Stephan Heucke.

The apex of such a cantilever's tip is roughly 20 nm in size. Assuming an average protein size in the order of 10 nm [68], only a few proteins are expected to be immobilized on the tip. This is important to ensure single-molecule interactions during an SMFS experiment.
In a typical SMFS experiment, the cantilever is slowly approached manually to its starting position, very close (low micrometer range) to the surface. When the tip of the cantilever comes in close proximity to the surface, the thermal oscillations of the lever are dampened, and therefore, the resonance frequency of the cantilever decreases. The resonance peak of a free BioLever mini cantilever is slightly above 20 kHz, and not detectable by the human ear. By listening to the thermal noise on the cantilever's deflection signal, the experimenter can carefully approach the cantilever and detect surface proximity: When the resonance frequency shifts to lower frequencies, the oscillation's resonance peak moves into the audible range and the signal appears louder. From this starting position on, a piezoelectric actuator moves the cantilever towards the surface and pushes into it until a preset force is reached. Receptor and ligand can form a complex, which upon retraction will be loaded mechanically. Upon loading of the complex, the cantilever is bent proportionally to the acting force. The laser beam, that is focused on its reflective back side gets deflected, resulting in a differential bending-dependent signal on a four quadrant photodiode (**Figure 5 A**).

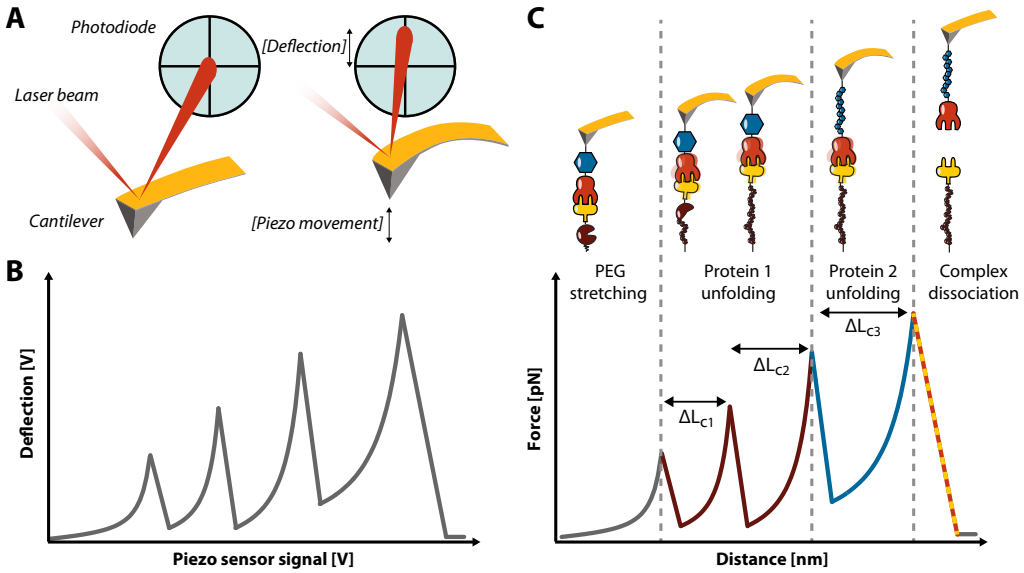


Figure 5. Schematic of a typical AFM SMFS experiment. **A** The AFM signal is detected *via* a four quadrant photodiode. After a complex between molecules on surface and cantilever has formed, the retraction of the cantilever mechanically loads the complex. The cantilever bends and deflects the laser beam. **B** Exemplary raw data trace of an AFM experiment. Recorded deflection is plotted against the position sensor signal of the piezoelectric actuator. **C** A typical force-distance diagram after transformation of raw voltage data into forces acting on the cantilever and distances of the cantilever tip to the surface is shown. First the linker molecules, usually PEG spacers on both sides (cantilever and surface), are stretched by retracting the cantilever. Then protein domains unfold hierarchically depending on their stability, the complex ruptures last. Adapted from [12] with permission from Elsevier. Copyright 2017, Elsevier.

The change of deflection is recorded by a four quadrant photodiode and converted into force with the spring constant of the force sensor and the optical sensitivity of the instrument. Piezo position z is recorded in nanometer by its capacitive sensor.

The cantilever acts as a spring that is assumed to be linear, hence Hooke's law can be employed [69]:

$$F = kz \tag{1}$$

With: k: Spring constant [pN nm^{-1}], z: Piezo position [nm]

The spring constant k can be calibrated based on the thermal fluctuations of the cantilever. The equipartition theorem applies since at thermal equilibrium, energies of oscillatory motion are equally distributed among its degrees of freedom [70]. Hutter and Bechhoefer found that a typical AFM cantilever (spring constant: $0.05\,\mathrm{N\,m^{-1}}$) with thermal fluctuations of 3 Å can be satisfyingly approximated by a simple harmonic oscillator with only one degree of freedom. Calibration of the piezo extension sensitivity is performed interferometrically by reflecting part of the deflection laser beam on a gold surface and measuring the change in sum signal on the four quadrant photodiode during a defined piezo movement. With knowledge of the laser wavelength, the piezo sensitivity can be derived from the recorded signal. These calibration values allow conversion of the deflection and position voltage signals into accurate force F and distance x values. Whenever the force is sufficient to let the complex overcome its thermal energy barrier for unbinding, receptor and ligand dissociate, and the cantilever relaxes into its initial state (**Figure 5 B**).
In order to verify a specific single-molecule interaction, a protein domain with known unfolding characteristics (unfolding force distribution, possible sub-steps, added contour lengths), a so called fingerprint domain is fused to the receptor and ligand domain individually.
In general, two types of SMFS experiments are possible: On one hand, a protein domain of interest can be probed by replacing one of the fingerprint domains. Ideally, the domain on the surface is chosen, especially when no prior knowledge about its refolding capabilities is given. On the other hand, it is possible to probe the mechanostability of receptor:ligand domains. This setup requires to choose well-known fingerprint domains with sufficiently low unfolding forces. Otherwise the complex ruptures frequently at forces lower than the fingerprint unfolding, and so control over the specificity is lost. **Figure 5 C** illustrates a typical AFM experiment: The linker molecules are stretched first and then the complex as well as the fingerprint domains are loaded with force simultaneously. Statistically, the weakest component in the chain unfolds first, in **Figure 5 C** "Protein 1", a fingerprint domain unfolds in two substeps. The two derived increments yield the exact length of the unfolded amino acid chain, known from previous experiments. Shortly after domain unfolding, the force drops nearly to zero, since the loaded fold relaxes into an unstructured and flexible amino acid chain, releasing additional free contour length. Next, the second weakest protein unfolds, in this case "Protein 2", yielding another increment in contour length. Now, that all fingerprint domains are unfolded, the complex is loaded with force until it ruptures. A new molecule can be probed, as long as the protein on the cantilever is not irreversibly damaged or unfolded. This is an essential requirement for a protein attached to a cantilever. The receptor domain and the fingerprint domain ideally refold quickly after unfolding or unbinding.
With such an experiment, the force necessary to dissociate a complex or unfold a protein domain at a given loading rate is obtained. In order to extract the contour length from a force-extension recording, various theoretical models describing the polymer elasticity have been developed. The worm-like chain (WLC) model describes the stretching of a polymer, and can be imagined that a protein backbone approximates the protein fold as a flexible polymer chain. The WLC model is one of the most commonly employed models in force spectroscopy (**Figure 6**) [69], [71], [72].

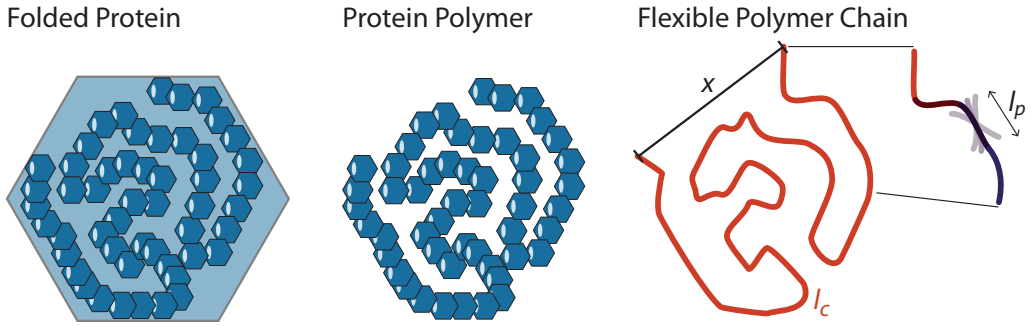


Figure 6. Illustration of WLC model. The WLC model describes the protein polymer as a flexible polymer chain. The contour length (l_c) is the length along the polymer backbone. x describes the-end-to end distance of N to C-Terminus, and the persistence length (l_p) describes the stiffness of the system. The smaller the persistence length, the more flexible the system.

The flexible polymer chain length is the contour length (l_c), which represents the whole protein backbone length. The persistence length l_p describes the smallest distance between two points along the protein backbone, at which directional correlation drops to e^{-1} [73]. For a given protein backbone, l_p is not necessarily uniform, because varying amino acid side chains influence the flexibility of the peptide backbone differently. For example smaller amino acids like alanine can move more unrestricted than bulky amino acids like phenylalanine. Hence, the persistence length of a polyalanine sequence is smaller than the polyphenylalanine sequence, despite both sequences consisting of peptide bonds. In a site-specific AFM SMFS experiment, the molecule is stretched and pulled apart from both termini. With unfolding of domains free contour length (already unfolded protein backbone) converges more and more towards its total contour length, since l_c is "hidden" in the protein fold and unravels during an unfolding experiment. Sub-domains under force unfold and stretch. The WLC model is approximated by [69], [71], [72]:

$$F(x) = \frac{k_B T}{l_p} \left(\frac{1}{4(1-\frac{x}{l_c})^2} - \frac{1}{4} + \frac{x}{l_c} \right) \tag{2}$$

With: F: Force [pN], k_B: Boltzmann constant: [$1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$], T: Temperature [K], l_p: Persistence length [nm], x: End-to-end distance [nm], l_c: Contour length [nm]

The WLC model sufficiently describes the stretching behavior of a protein in a low force regime up to approx. 100 pN, where mostly entropic contributions influence the system [74]. Higher forces may deform bond lengths and angles. The WLC model can be extended with an empirical, Hookean (specific) stiffness [69], [75], which accounts for additional stretching of the protein backbone, considering enthalpic and entropic contributions:

$$F(x) = \frac{k_B T}{l_p} \left(\frac{1}{4(1-\frac{x}{l_c})^2} - \frac{1}{4} + \frac{x}{l_c} \right) + \Phi x \tag{3}$$

With: Φ: Specific Stiffness [$\mathrm{pN\,nm^{-1}}$]

The extended WLC model describes protein unfolding behavior only sufficiently within forces up to 200 pN. Hugel *et al.* incorporated a quantum mechanical correction term, as an *ab-initio* alternative to the Hookean stiffness correction to account for backbone stretching [76].

Livadaru *et al*. established a dynamic model describing data for a larger force range [77].

$$\frac{x}{l_c} \simeq \begin{cases} \frac{Fa}{3k_BT} & for \quad \frac{Fb}{k_BT} < \frac{b}{l_p} \\ 1 - \left(\frac{4Fl_p}{k_BT}\right)^{-\frac{1}{2}} & for \quad \frac{b}{l_p} < \frac{Fb}{k_BT} < \frac{l_p}{b} \\ 1 - \left(\frac{2Fb}{k_BT}\right)^{-1} & for \quad \frac{l_p}{b} < \frac{Fb}{k_BT} \end{cases} \tag{4}$$

With: a: Kuhn length [nm], b: Segment length [nm]

With the help of introduced models protein unfolding can be described, deriving energies necessary to unbind or unfold complexes and gain insight into their molecular mechanics. **Figure 7** illustrates the difference between kinetic unfolding (red) and force-dependent unfolding (blue). While applying force to a system (F), the transition state from folded to unfolded state is energetically lowered. The higher the applied force, the lower the energy barrier becomes.

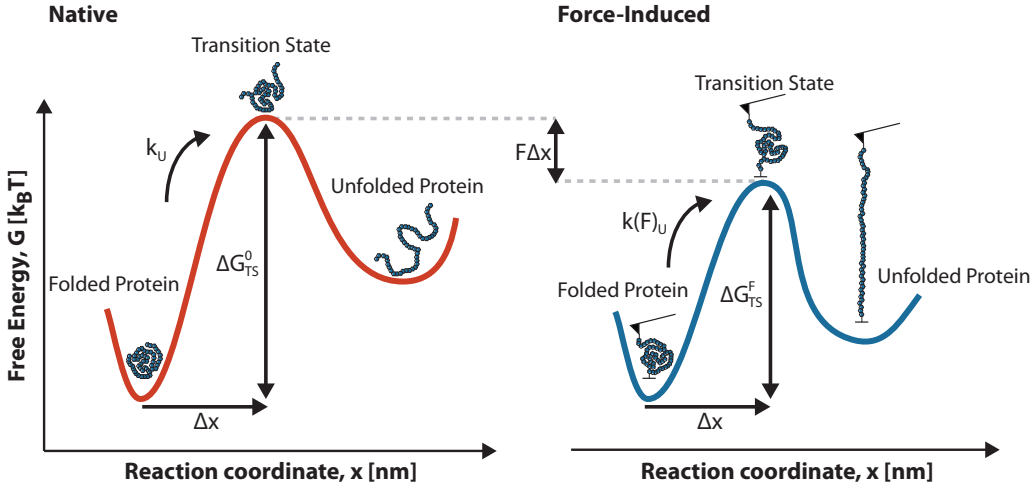


Figure 7. Schematic of the influence of force on a two-dimensional unfolding energy landscape. Folded proteins are thermally excited across a transition state with an unfolding rate k_U or $k(F)_U$ and ΔG, the energy necessary. During force F application, the energy barrier of the transition state is lowered, the unfolding is more probable in a given time frame. The lowering of the energy barrier of the transition state only happens in the force loaded path along x of the multi dimensional space of the energy landscape, all the other directions and unfolding possibilities might not necessarily be affected.

After an experiment, the obtained data traces can be analyzed with the appropriate models. The probabilities of certain unfolding or unbinding events are dependent on the loading rate r and can be described with $F^*(r)$ [69], [78]–[81]:

$$F^*(r) = \left(\frac{k_BT}{\Delta x}\right) ln\left(\frac{r\Delta x}{k_BTk_{off}}\right) \tag{5}$$

With: $F^*(r)$: Most probable rupture force [pN^{-1}], Δx: Distance of folded to unfolded state on energy landscape [nm], r: Loading rate [pN s^{-1}], k_{off}: Off-rate [s^{-1}]

Rupture forces generally increase with the loading rate r (and therefore with the pulling speed in constant speed protocols) [69], [82]. AFM experiments are typically conducted at timescales of 10^{-3} to 10^1 s, at which thermal fluctuations dominate the unbinding processes [69]. In the thermally driven region investigated with standard

AFM SMFS, the unbinding or unfolding force often scales linearly with the logarithm of the loading rate, according to the Bell-Evans model (see **Equation 5**) [80], [81]. The higher the loading rate in an experiment, the shorter the time scale for dissociation or unfolding. As a consequence, higher forces are necessary to lower the energy barrier sufficiently to an extent for thermal unbinding or dissociation to take place (**Figure 8**).

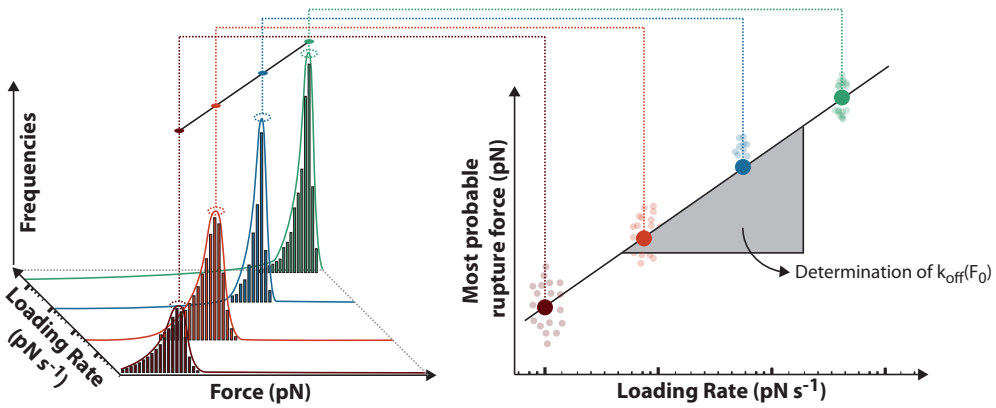


Figure 8. Illustration of the correlation of loading rate and force. Force scales logarithmically with increasing loading rate. At higher loading rates the time scale of an unfolding or dissociation event is shorter. To compensate the shortened time scale, the rupture force increases, and thereby the energy barrier lowers, increasing the probability to thermally overcome the barrier. Different pulling speeds within one experiment can be plotted in force-loading rate plot. Fitted parameters in this plot can be used to determine the theoretical off-rate at zero force for the AFM determined unfolding path.

Theoretically it is possible to derive the zero force off-rate from the force-loading rate dependency. However, this is only a single path on the energy landscape, biased by force applied *via* AFM [69], [83]. This off-rate at zero force under a single unbinding pathway does not necessarily reflect the bulk off-rate, which can be used to describe the affinity of a complex. The bulk off-rate consists of a combination of different unbinding pathways. The affinity of a receptor:ligand complex is defined by its ratio of unbinding and binding rate [69], [84].

$$K_D = \frac{k_{off}}{k_{on}} \tag{6}$$

With: k_{on}: On-rate [$\mathrm{M}^{-1}\,\mathrm{s}^{-1}$], K_D: Equilibrium constant [M]

Since the equilibrium constant K_D is concentration dependent, a single molecule complex is better described with its lifetime $\tau = (k_{off})^{-1}$ [69], [81], [85], [86].

$$\frac{1}{\tau(F_0)} = \omega \cdot e^{-(\frac{\Delta G^0_{TS}}{k_B T})} \tag{7}$$

With: $\tau(F_0)$: Lifetime at zero force [s], ω: Attempt frequency of a system [s^{-1}], ΔG^0_{TS}: Activation Energy Barrier [J].

In order to determine accurate, force-dependent lifetimes τ of a complex, more sophisticated experimental setups are necessary.
Typically, AFM experiments are performed in constant speed mode (constant movement of z-piezo), because the experimental setup is relatively simple [87]. In constant

speed mode force is usually plotted against extension to obtain information about mechanical stability of a fold.
The more sophisticated force-ramp mode (constant loading rate) requires the implementation of feedback loops in the experimental setup which is more difficult to realize [88]. Fast reacting feedback loops are necessary to keep the loading rate constant during an AFM experiment. If force is kept constant until an unfolding or unbinding event occurs, essentially force-ramp mode at loading rate of $0\,\mathrm{pN\,s^{-1}}$, it is called force-clamp mode. Force-clamp mode allows the determination of the lifetime of a protein fold or complex at a given force [89].
The dissociation process can be accelerated by increasing the systems temperature [90]. This also holds for a protein domain, an increase in temperature can be used to usually accelerate protein unfolding. An increase in temperature increases the fluctuations of the system and thus the attempt frequency to overcome the energy barrier to the unfolded state.
In contrast to applying force, which lowers the energy barrier of a system (see **Equation 8**). The higher the applied force the lower the average lifetime of a complex $\tau(F)$. This correlation allows the derivation of a force-dependent lifetime $\tau(F)$, that decreases exponentially with force [69], [91]:

$$\frac{1}{\tau(F)} = \omega \cdot e^{-(\frac{\Delta G^0_{TS} - F\Delta x}{k_B T})} \tag{8}$$

Equation 7 and **Equation 8** can be simplified to [69], [92]:

$$\tau(F) = \tau(F_0) \cdot e^{-(\frac{F\Delta x}{k_B T})} \tag{9}$$

Similar to force-loading rate plots, force-lifetime plots can be obtained by varying the clamping force in a force-clamp experiment. A fit through all lifetime populations allows extrapolation of the lifetime at zero force for that specific unfolding path.
Both methods, constant force and force ramp, are able to decipher molecular mechanisms in protein folds and receptor:ligand interactions.

2.4 AFM Data Analysis

Thousands of force-distance traces are recorded, during a AFM experiment. Only a fraction contains single molecule curves with both fingerprint domain unfoldings and final complex dissociation (**Figure 9 A and B**). Panel B shows a so called shielded event in the final rupture. Usually, domains are likely to unfold sequentially from weakest to strongest fold. Sometimes, a stronger domain topologically shields a weaker (sub-)domain, which can only unfold after the stronger domain is unfolded and is therefore termed shielded event. In the other recorded traces either no interaction between receptor and ligand forms (**Figure 9 C**), or more than one complex was formed, yielding non-usable multiple interaction traces (**Figure 9 D**), non-specific interactions between cantilever and surface formed (**Figure 9 E**), or proteins did already partially unfold before probing or do not unfold completely while probing (**Figure 9 F**). It is necessary to filter all these unwanted curves (red traces) manually or *via* algorithm sorting for fingerprint domain unfolding and receptor:ligand rupture signatures. An automated routine can sort and analyze all suitable traces, for example to identify protein folds, measure force dependencies or measure mechanical stability.

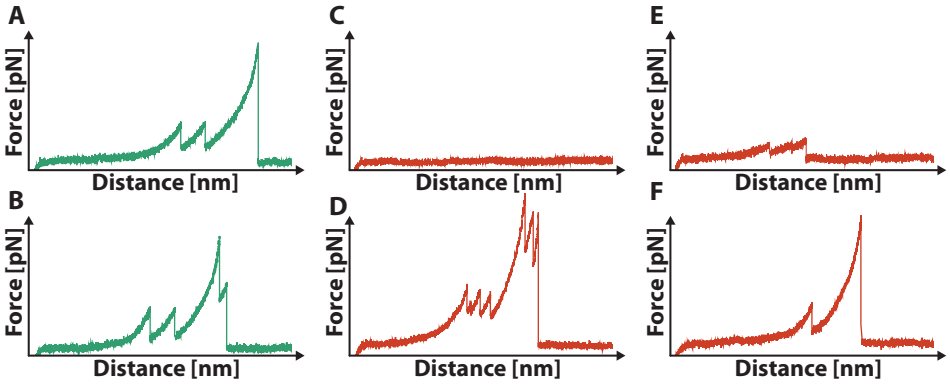


Figure 9. Overview of typical traces in an SMFS experiment. During an AFM experiment different kinds of traces are obtained. Green are curves usually viable for further data analysis (A/B). Red depicts non-usable curves for data analysis (C-F). Characteristic curves are: **A:** A single receptor:ligand interaction with two fingerprint domains unfolding, **B:** A shielded single receptor:ligand interaction with two fingerprint domains unfolding, **C:** A flat line without any interaction, **D:** Multiple interactions between more than one receptor and ligand, **E:** A non-specific interaction between cantilever and surface, **F:** An incomplete single receptor:ligand interaction (here: only one fingerprint domain unfolded).

After sorting, usually 0.1 - 20 % of the curves yield clear single molecule interactions with fingerprint unfolding events on both sides (**Figure 10 I**). The reduced data set with all specific single interactions can then be analyzed in two alternative ways.
In the first approach, all curves are transformed from force-distance space into force-contour length space (**Figure 10 IIa**) [93]. Contour length histograms are assembled. An adequate bin size of the histograms usually leads to sharp distributions, which can be cross-correlated (**Figure 10 IIIa**). A random curve is selected as initial template and cross-correlated to another one on the contour length axis. Each curve is offset for their maximum correlation value. Both aligned histograms are now added and used as template to align the next curve. This procedure is repeated until all curves are aligned. Finally, the outcome of the cross-correlation is aligned again against each curve of the set, to reduce bias and referencing effects depending on the initial choice

of curve. In the end, a probability density can be plotted for the whole experiment, representing the most likely energy barrier positions given through the unfolding events (**Figure 10 IV**) [94].
An alternative way of analyzing the data is to assemble a representative master curve in force-distance space, formed from all specific unfolding traces. Here, single force-distance traces are cross-correlated, similar to the method described above. However, instead of transforming into contour length space, the traces are aligned in the force-distance space (**Figure 10 IIb**) yielding a "master curve". This allows direct data handling in the force-distance space, like fitting of WLC models to the traces to describe the polymer elasticity and unfolding of domains. It is also possible to transform the master curve into contour length space to determine energy barrier positions. The master curve (**Figure 10 IIIb**) reflects the most probable pathway of all curves from one experiment.

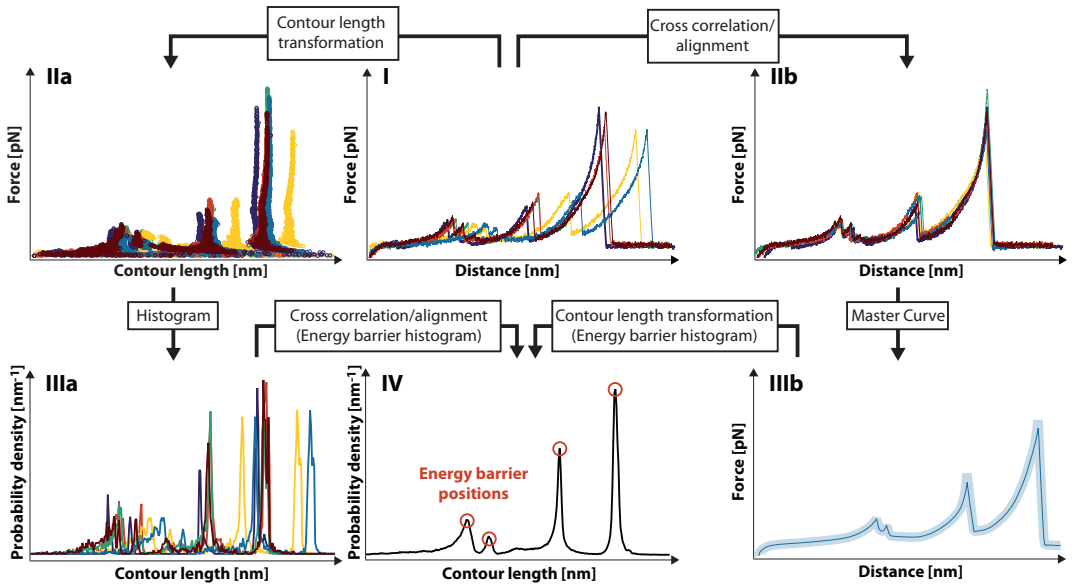


Figure 10. Schematic of the data processing of AFM force traces. There are two methods of analyzing the raw data. (**Ia)** In the first method, raw traces are transformed into contour length space (**IIa**) and histograms with nanometer bin size are plotted (**IIIa**). Then, the histograms are cross-correlated and aligned to obtain an overlay of all curves (**IV**). In the second method, data traces in the force-distance space are cross-correlated (**IIb**). A master curve of the aligned force-distance trace represents the most probable unfolding trace of this set of curves (**IIIb**). As in the first method the master curve can be transformed into contour length space and an energy barrier histogram is obtained (**IV**). Adapted from [12] with permission from Elsevier. Copyright 2017, Elsevier.

Independent of the method of data analysis, most probable rupture or dissociation forces are obtained and can be used to address questions about the mechanobiolgy of a complex or a protein fold.

2.5 Linkers in SMFS

Linkers are crucial in SMFS studies. They provide proper passivation against non-specific adsorption of the biomolecules to the surface [95]. Non-specific adsorption is undesirable as molecules will be picked up by the cantilever, blocking cantilever molecules bonding partners: Non-specifically adsorbed proteins are not covalently anchored to the surface, hence they might not be removed from the cantilever upon retraction. Once all binding domains on the cantilever side are saturated with non-specifically adhered proteins, no new complexes can be formed, thus no new traces are recorded - the experiment stalls until the complex dissociates naturally. However, non-specifically adsorbed proteins on the cantilever are no problem since there is an excess of surface-anchored proteins.
Most importantly, linker molecules provide spacing between the surface and the biomolecule of interest. Otherwise proteins might interact with the surface and create undesired unfolding artifacts, *e.g.* through partial adsorption to the surface.
The anchoring point of the linker molecule, especially on the cantilever, also influences the experiment [96]. Different attachment sites lead to different unfolding pulling angles and might result in varying unfolding forces and pathways, biasing the rupture force distribution of a protein or complex.
The most commonly employed linker so far is polyethylene glycol (PEG). This polymer is commercially available with different reactive groups and in different molecular weights and therefore lengths. However, with the advancement of SMFS to access higher forces, new challenges appeared: PEG has a force-dependent conformational change (**Figure 11**).

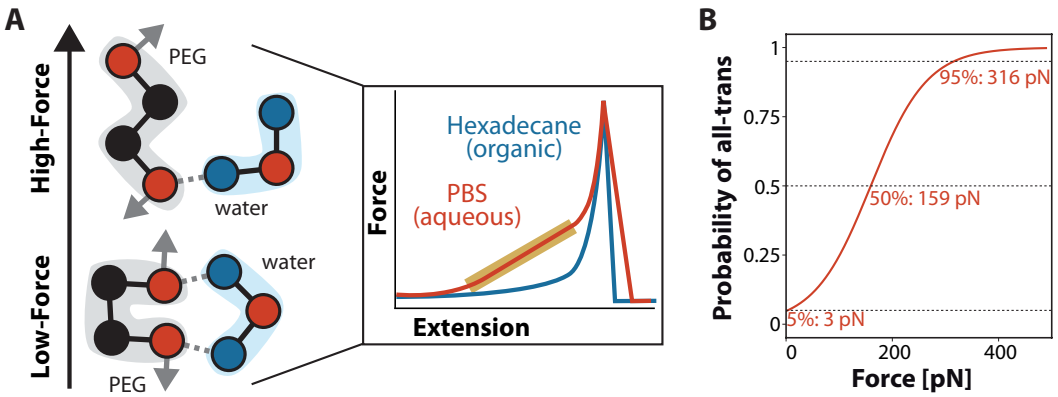


Figure 11. Schematic of the force-induced conformational change of PEG. A PEG is usually stabilized in water by hydrogen bonds in a trans-trans-gauche conformation. Increasing force shifts the equilibrium towards an all-trans state. The inset in the middle shows the behavior only exists in aqueous systems (Red trace, in PBS buffer), in organic solvents hydrogen bonds do not exist (Blue trace) [97]. Yellow highlights the linear regime of the force-induced conformation transformation. **B** The force-dependent distribution probability of PEG and its all-trans conformation state. Adapted from [97], under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0).

This phenomenon was first described by Oesterhelt *et al.* in SMFS [97]. In aqueous environments, which all buffer systems for biomolecules are, PEG forms hydrogen bonds with water molecules. The oxygen atoms of the PEG backbone interact with hydrogen atoms of water, leading to a more compact conformation. Force increases the probability of a conformational switch. With the robust receptor:ligand pair

from type III cohesin:dockerin a high force regime is probed in an experiment (over 316 pN). Based on Oesterhelt's theoretical framework the probability for the change from trans-trans-gauche to all-trans conformation at 159 pN is 50 %. At forces of 316 pN the probability of PEG being in all-trans conformation is about 95 %. Hence, in an experiment with elevated forces the PEG conformation change leads to a net increase of the contour length, which complicates data analysis.
So, Tong *et al.* suggested alternative chemical linkers for force spectroscopy [98]. Instead of using PEG they synthesized a phosphoramidate-based linker. They showed the feasibility of the linkers for DNA/DNA-, DNA/Protein- and Protein/Protein-interactions. Furthermore they were able to produce monodisperse linkers, compared to polydisperse PEG, leading to a more consistently functionalized surface. The molecule which is linked closest to the tip apex forms complexes with the receptors on the surface with the highest probability. Polydispersity in linkers broadens the area with molecules available for complex formation and increases the binding probability of different molecules. Producing these monodisperse, phosphoramidate-linkers solves that issue, however due to their chemical, solid state-based synthesis they require additional coupling steps to DNA, peptides or proteins.
With the new possibilities in protein conjugation techniques and DNA synthesis it is now possible to employ biological linkers in SMFS [99]. In order to address that issue Ott and Jobst *et al.* proposed to replace PEG with new recombinantly produced biological linkers - Elastin-like polypeptides (ELPs) [100]. ELPs are biological polymers with a temperature-dependent cloud point: Exceeding a certain temperature ELPs change their conformation and aggregate, which can be reversed by lowering the temperature again. The cloud point is dependent on the amino acid composition of the repeating pentapeptide sequence "VPGXG", with X being any amino acid except proline (**Figure 12**). The inclusion of basic or acidic amino acids render the polymer pH sensitive. Other factors influencing the cloud point are salt concentration or molecular weight of the ELP, making them straight forward to purify by their unique properties [101].

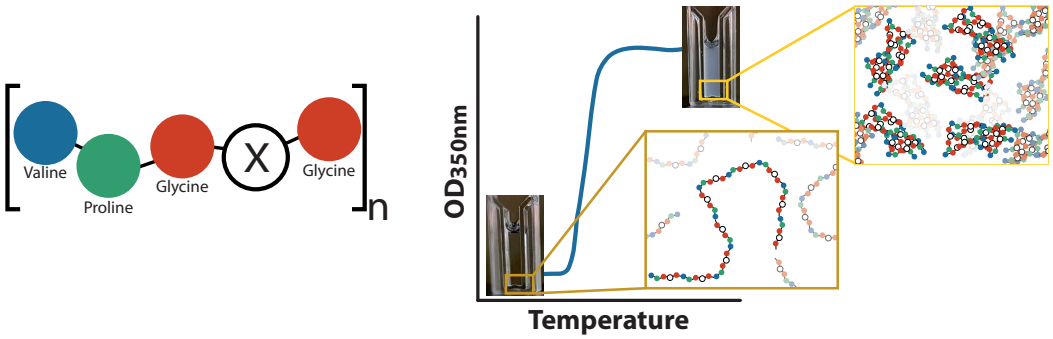


Figure 12. Illustration of Elastin-Like Polypeptides. ELPs are composed of repeats of the pentapeptide Valine-Proline-Glycine-X-Glycine. X represents any amino acid except proline. ELPs collapse at a certain temperature and precipitate in solution, which can be measured by absorption at 350 nm. The precipitation is reversible, lowering in temperature leads to resolubilization of the ELPs and the solution becomes clear again. The yellow boxes illustrate the process: ELPs change their conformation and interact with each other, which leads to larger aggregates clouding the solution.

The benefit of smart polymers like ELPs is that they can be modified according to the needs of the experiments as they can be produced recombinantly with different conjugation tags already in place. Furthermore they can be easily purified with the

inverse temperature cycling (ITC) method [102]. ITC precipitates and redissolves ELPs by heating and cooling, whereas other proteins, *i.e.* contaminating host proteins, remain precipitated. After a few iterations only pure ELPs are left in solution.
Like phosphoramidate linkers, and in contrast to PEG linkers, ELPs are monodisperse an additional advantage to SMFS. Monodisperse and defined linkers minimize measuring artifacts and increase the robustness of the system, contributing to reliability and ease of analysis of obtained data.

2.6 Surface Chemistry in Single Molecule Force Experiments

With the advances in surface chemistries and site-specific attachment strategies different pulling geometries can be probed. Adding a tag for covalent surface chemistries at either N- or C-terminus orients molecules correspondingly. In total, four different full-length pulling geometries can be probed to gain insights into native and non-native tethering geometries (**Figure 13**).

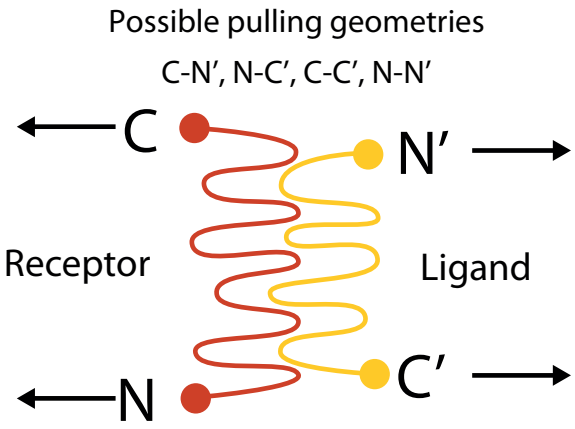


Figure 13. Site-specific attachment sites and binding geometries. With site-specific attachment the (un)binding geometry of complexes and proteins can be defined specifically. Red illustrates the receptor, yellow the ligand. In total four different geometries can be probed. Reprinted from [12] with permission from Elsevier. Copyright 2017, Elsevier.

The tethering geometries of a molecule has an impact on its stability. Schoeler *et al.*, showed that the type III cohesin:dockerin interaction depends on the orientation of the cohesin. When the cohesin domain is located at the N-terminus rupture forces up to 650 pN were achieved. When located at the C-terminus the complex already ruptured at forces of 150 pN [13], [14]. Hence, it is necessary to anchor proteins site-specifically to be able to differentiate between different unfolding or dissociation forces.

Advancements in DNA synthesis, bioconjugation techniques and protein expression and purification strategies enable a fast throughput in protein generation. Depending on the requirements of the assays, it is possible to synthesize the same protein with tags in different locations.

By now, several reliable covalent immobilization strategies are well established (**Figure 14**) [12]. Most methods rely on certain reactive groups of different amino acid side chains, *i.e.* amines from lysines or carboxyl-groups from glutamates, as well as the N- and C-terminal amine- and carboxyl-groups to immobilize proteins. However, most proteins contain more than one accessible, reactive residue, which results in heterogeneous surface pull-down and thus orientation of molecules.

The less present cysteine is an attractive alternative to the employed amine- or carboxyl-groups. Cysteine can be covalently linked to maleimide-groups *via* their free thiol-group. However, it is only applicable when no other cysteine is accessible and the protein needs to be insensitive against reduction agents. Reduction before immobilization is necessary to break intermolecular disulfide bond formed during protein production and purification.

A more robust and versatile approach is to rely on enzyme-mediated pull-down strategies. Several enzymes were characterized and optimized during the last decades

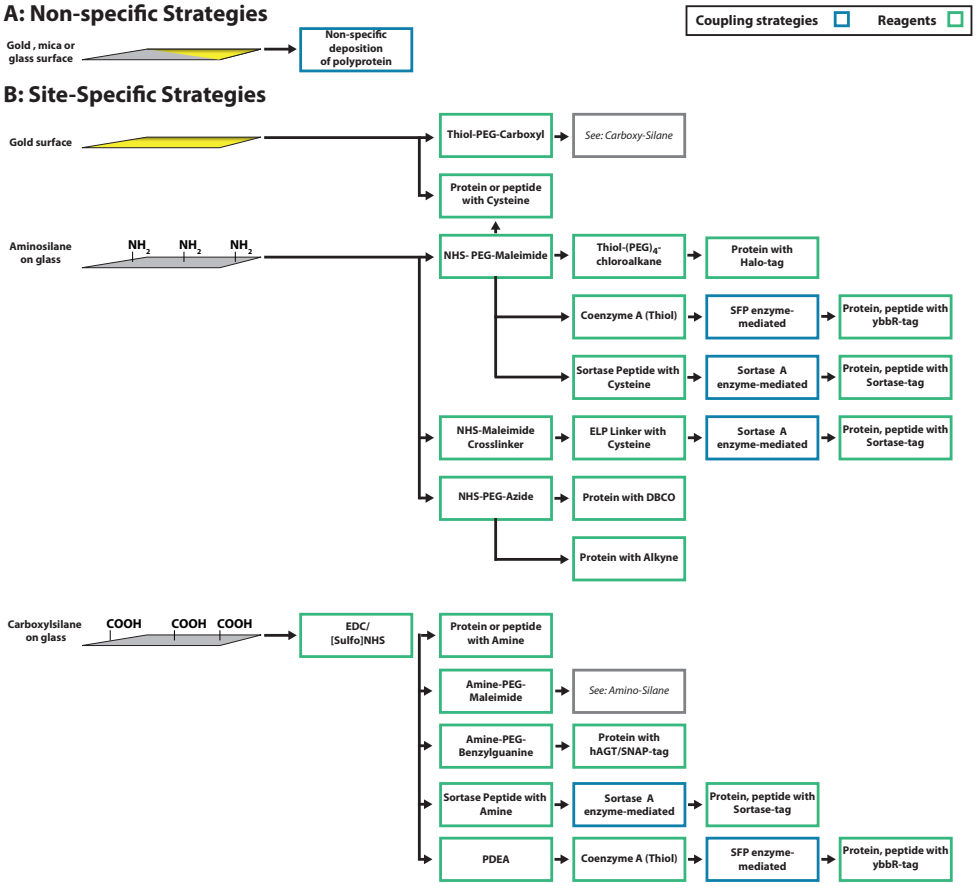


Figure 14. Schematic of possible protein immobilization strategies. Depending on the surface material (gold or glass) and its surface functionalization (amino- or carboxylsilane) different surface coupling strategies can be employed. It can be distinguished between non-specific and specific attachment. Coupling strategies are labeled in blue, employed reagents in green. **A** The non-specific approaches rely on the interaction between charged residues of surface and molecules to be deposited. **B** Amine-, carboxyl- or thiol-groups of proteins can be used to covalently link proteins to activated groups on surfaces. The top row describes amine-thiol crossreacting strategies. In this case the amine-group of the aminosilane is crosslinked with NHS-groups, which is most often linked to maleimide-groups. Maleimide-groups react with thiol-groups. A new approach represents the incorporation of non natural amino acids. They are a bio-orthogonal chemistry and can be used to pull-down proteins internally. The lower row describes strategies to crossreact carboxyl-groups with amine-groups. It is possible to crosslink carboxylsilane with EDC/NHS and functionalize the glass surface with amine-groups, *i.e.* amine-benzylguanine-PEG, necessary for hAGT or SNAP-Tag technologies. Adapted from [12] with permission from Elsevier. Copyright 2017, Elsevier.

[12]. All these enzymes catalyze the covalent coupling of a small molecule or peptide to a protein.

Two commonly used surface coupling strategies in force spectroscopy are enzyme-based (Sfp and Sortase A) [12]. The easiest way to employ the two enzymes, is to start with functionalization of surfaces and cantilever with aminosilane. The amine-groups are subsequently crosslinked with NHS-maleimide, offering thiol-reactive surfaces and cantilevers suitable for subsequent enzyme-mediated immobilization [103].

Sfp is a magnesium-dependent 4'-phosphopantetheinyl transferase from *Bacillus subtilis* and couples Coenzyme A (CoA) to a hydroxyl-group of a serine of an 11

amino acid long recognition sequence (ybbR-tag) [104]. CoA contains a thiol-group, which is not important for the Sfp catalysis step, such that it can be used for coupling to a maleimide-group.
The sortase A, from *Staphylococcus aureus* is an calcium-dependent peptidase, and has been optimized for post-translational protein coupling strategies [105]. It ligates N-terminal glycine residues (at least one) to a C-terminal LPETGG-peptide sequence. For SMFS experiments a cysteine containing peptide with the LPETGG- or the GGG-motif is coupled, depending on the orientation of the protein of interest, to the maleimide-activated surface. During the reaction the C-terminal GG is cleaved and the N-terminal glycines are linked to the C-terminus *via* a peptide bond, restoring the LPETGG-motif. This means the product can be cleaved again if educt with N-terminal glycine is still present or added.
Using Sortase A can be advantageous in force spectroscopy because it offers the possibility of *in situ* protein ligation. Proteins can be produced with only a very small N- or C-terminal Sortase-tag and anchored with an orthogonal tag on the opposite terminus, *e.g.* a ybbR-tag. Small peptide tags are presumably minimal invasive, *i.e.* do not influence the native fold of the protein and do not lower expression yield in *E. coli* [106]. After surface functionalization with Sfp *via* ybbR-tag, the protein of interest can be modified with a receptor domain *via* Sortase A, to be used as a binding handle.
The SpyTag-/-Catcher system is a relatively new development in biochemistry. SpyCatcher is a domain which recognizes the SpyTag, a short peptide sequence, and spontaneously forms an isopeptide bond with it after binding. Hence, proteins can be anchored to SpyCatcher-activated surfaces without addition of external, catalyzing enzymes or reagents [107].
Whenever the mechanobiology of subdomains of a protein fold is subject of a study it is necessary to immobilize the protein internally. Only internal immobilization allows the direct stretching of an isolated subdomain, without biasing effect of surrounding folds. Internal immobilization is not possible with Sortase A, since it cleaves parts of the protein before ligation. Sfp- and SpyCatcher-based approaches are also not optimal, as their recognition sequence are several amino acids long and might interfere the overall fold of the protein. These approaches can only be used at flexible loops which do not contribute to protein folding and are accessible for Sfp or SpyCatcher.
Introducing an internal cysteine is one way to anchor proteins internally in force spectroscopy experiments [108]. However, this method only works reliably if only one accessible cysteine is present. Otherwise multiple geometries will be probed during an experiment.
The development of non natural amino acid pull-down strategies, especially alkyne and azide containing side groups, are a viable alternative. They are inert in biologically relevant reactions, *i.e.* cannot be inactivated during protein expression and do not require any reduction steps [109].
Independent of the site-specific immobilization strategy, in an experiment chosen proteins should be covalently and site-specifically attached to linkers at the apex of a cantilever tip or the glass surface, so their biomechanics can be probed unambiguously.

2.7 Biomolecule Preparation for SMFS Experiments

Developments in DNA cloning techniques allow to combine desired DNA coding sequences for a protein, similar to assembling building bricks [11]. This facilitates production of proteins for SMFS studies. The protein of interest needs to be tagged and modified in several ways for SMFS experiments: A purification-tag, an immobilization-tag for site-specific attachment to surfaces and optionally a fingerprint domain and a binding handle from a receptor:ligand system.
The immobilization- and purification-tags are usually attached at the N- or C-terminus of the fingerprint domain, sufficiently spaced so that they do not interfere with complex formation in the SMFS experiment.
DNA assembly methods are needed to combine all these coding sequences. **Figure 15** depicts the basic principle of two established cloning methods for DNA preparation in SMFS studies.

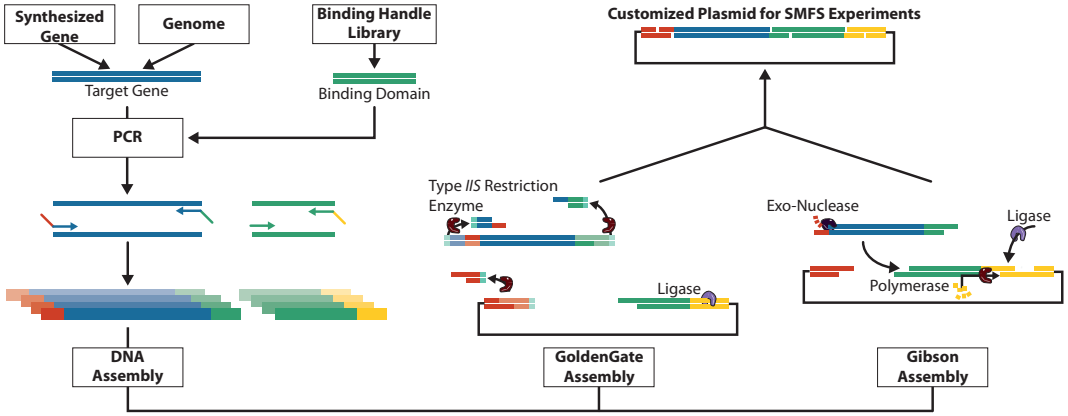


Figure 15. Schematic of two viable DNA assembly methods for SMFS experiments. Modern DNA cloning methods allow scarless assembly of different DNA fragments. A typcial SMFS construct contains a purification- and immobilization tag, a fingerprint domain and a binding domain. Depending on the scientific question, purification and immobilization strategy different suitable domains can be assembled. DNA is amplified with customized primers in a PCR reaction. Primers are designed to leave overlaps to neighboring DNA fragments. Templates for the reaction are synthesized genes or genomes from living organisms. The different PCR products, encoding necessary genes and domains are hierarchically assembled with either GoldenGate or Gibson Assembly. In the end a scarless fusion of different domains is ligated and subcloned into a vector.

Either genes are synthesized chemically or DNA from genomes or gene fragments of interest are amplified *via* polymerase chain reaction (PCR). Depending on the assembly method (GoldenGate [110] or Gibson Assembly [111]), primers need to be designed correspondingly. Both methods are scarless, *i.e.* they generate no undesired coding sequence that might change the native protein in the assembly process, which is essential for single molecule studies.
GoldenGate relies on restriction digest facilitated by type *IIS* restriction enzymes, which cut next to their recognition sequence [110]. Hence, primers can be designed with their recognition site next to their annealing site. Upon DNA-cleavage by the enzyme, the recognition site is removed and leaves sticky four base pair (bp) sequence that overlaps with the adjacent part. A ligase, ideally one which does not ligate blunt DNA-ends (*i.e.* T7 ligase), links the overlapping DNA fragments covalently. The

short overlaps only consist of four bp, which makes GoldenGate approach suitable for cloning of highly repetitive DNA.
Gibson Assembly relies on three enzymes working under isothermal reaction conditions (50 °C) [111]. A heat-labile T5 exo-nuclease digests DNA from the 5′ to 3′ end of double stranded DNA, generating DNA overlaps of several bp that can anneal with other overlapping DNA fragments. The T5 exo-nuclease degenerates over time, stopping it from fully digesting one DNA strand. The polymerase fills the excised nucleotides (5′ to 3′ end) after annealing of single stranded overlaps. The ligase covalently links the annealed fragments in parallel. The method relies on unique overlapping regions, complicating the cloning of highly repetitive proteins, as alignment possibilities are not unique anymore. Unlike GoldenGate assembly, repetitive motifs are hard to clone because the overlapping region is more than four bp long. However, Gibson Assembly is convenient in generating libraries, *i.e.* plasmids with different fingerprints from different templates since it is not required to delete any undesired restriction sites prior to gene amplification.
The plasmid (**Figure 16**) can now be assembled. Additionally a plasmid contains regulatory elements, *i.e.* antibiotic resistance gene to maintain the plasmid in cells. The origin of replication is necessary for the plasmid multiplication. A lac repressor enables the induction of protein expression.

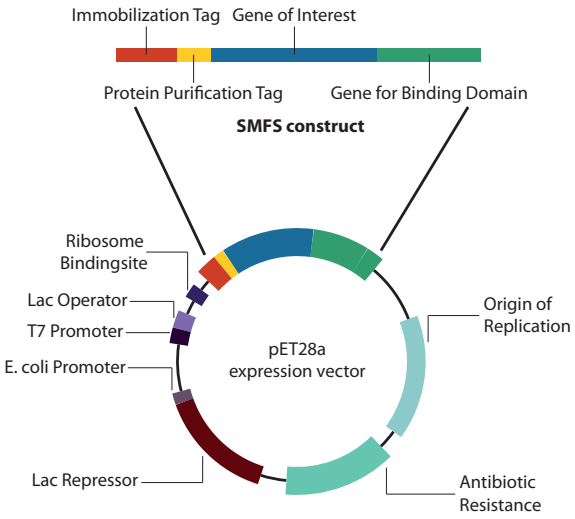


Figure 16. Illustration of a pET28a expression vector used for protein biosynthesis. A typical procaryotic expression vector contains different regulatory elements besides the gene of interest. A repressor enables induction of protein synthesis at a given point with a certain inducer molecule. In the case of pET28a the lac repressor (dark red) attaches to the lac operator and inhibits mRNA synthesis of the gene of interest until lactose or IPTG is supplied. The origin of replication (light cyan) ensures plasmid amplification in *E. coli* cells. The selection marker (antibiotic resistance) is needed to apply selection pressure to maintain the plasmid. A typical construct used in SMFS contains a immobilization tag, a folded (fingerprint) domain, a binding domain and a purification tag.

Depending on the assays and the protein of interest two expression methods are suitable: The classical approach with recombinant expression in a host cell or the cell-free approach *in vitro*. **Figure 17** sketches the principles of protein expression *in vivo* and *in vitro*.
The classical approach relies on recombinant protein production within host cells. Most commonly *E. coli* is transformed with the plasmid for protein expression, but

also yeast or insect cells are suitable.

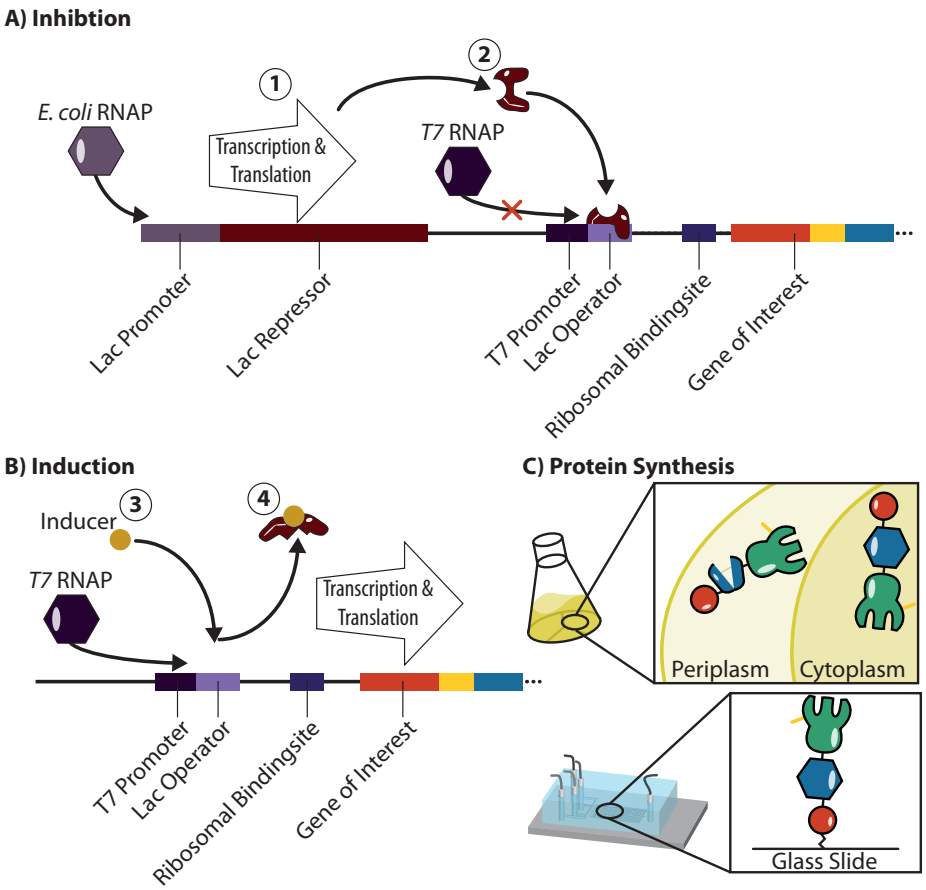


Figure 17. Sketch of protein expression principles and their application in cell-based and cell-free environments. A *In vivo* protein expression it is necessary to inhibit protein expression initially, because of the enormous metabolic load during expression, thus securing fast growth to higher cell densities. This is facilitated through binding of the basally expressed lac repressor to the lac operator (1). It blocks the T7 RNA polymerase (RNAP) from synthesis of the target mRNA (2). **B** At high cell densities an inducer is added to the media (3) that releases the lac repressor (4) from the lac operator and enabling mRNA synthesis through T7 RNAP, ensuring high protein expression yields **C** With cell-based protein synthesis strategies it is possible to either produce proteins in the cytoplasmic or periplasmic space. This has an impact on protein folding and formation of disulfide bonds, preferably generated in the oxidizing environment of the periplasm. During a cell-free synthesis this is not possible. However, proteins can be synthesized and immobilized in a one pot reaction.

In *E. coli* cells transcription of the used expression cassette is repressed in its basic state [112]. This is ensured by the continuous expression of the lac repressor (on plasmid) by an endogenous RNA polymerase (RNAP). The lac repressor binds to the lac operator and sterically hinders the genomically encoded T7 RNAP to synthesize the mRNA of the gene of interest (**Figure 17 A**). The exogenous T7 RNAP is used for synthesis of the target mRNA, to minimally interfere with the metabolism of the host cell [113]–[115]. At a defined point, an inducer molecule is added to the media which starts protein expression by inducing conformational change in the repressor and leading to its detachment from the lac operator (**Figure 17 B**). Now, the T7 RNAP can transcribe the mRNA and ribosomes translate the mRNA to protein.

Different expression and growth media have been developed over the years, employing the lac repressor/operator system for induction. Classically protein expression is induced *via* addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside), a lactose derivate, which *E. coli* cannot metabolize. Alternatively modern complex growth media contain a sugar mix consisting of glucose, glycerol and lactose [116]. *E. coli* prefers glucose over lactose as carbon source [117], only if all glucose has been consumed it switches to its lactose metabolism. This has the advantage that cells can grow to high densities initially upon glucose uptake and with its consumption they switch to lactose, hence induce protein production automatically at a later stage in their growth phase.
Subsequently cells are lysed which releases the proteins of interest. These can be isolated from the cell debris and purified with their purification-tag. The two most established purification systems are a HIS-tag (six repeats of histidine) or the StrepII-tag (eight amino acid tag). The HIS-tag binds to nickel, which is usually complexed with Nitrilotriacetic acid (NTA) bound to beads [118]. Streptactin binds the StrepII-tag, which can also be fused to beads [119]. The beads can be filled into a column, enabling a streamlined purification process. The cell lysate is applied to the column, the protein of interest binds to the material, host contaminants are flushed trough and washed away. Upon administration of an elution solution the pure protein of interest is released and can be either stored for later use or directly immobilized and probed in SMFS experiments.
Cell-free protein synthesis facilitate a faster throughput, since a plasmid with essential components for living cells is not strictly required. Hence, a linear gene cassette obtained from the DNA assembly is sufficient to produce proteins. The expression reaction can be directly initiated without the need of DNA amplification in *E. coli*. Additionally, cell-free approaches do not rely on induction because there is no need to balance the metabolic load [120], [121].
Preparing linear DNA and applying it to a one pot reaction (biosynthesis of proteins and immobilization reaction in parallel) is faster than the classical way of preparing proteins and especially attractive for multiplexed SMFS studies. In addition to the accelerated generation of biomolecules, different expression systems can be employed: *i.e.* procaryotic, eucaryotic, plant or complete synthetic reaction mixes [120], [122], [123]. Toxic proteins can also be expressed [124].
However, cell-free approaches can only produce proteins for a short amount of time until they run out of resources. Also, large proteins, proteins forming disulfide bridges, complex folds dependent on chaperone proteins or secretion, still need to be produced in *E. coli*.
Independent of the protein production approach, constructs for SMFS contain usually the immobilization-tag on the opposite end from the binding domain. This ensures that only full-length constructs are probed in the experiments. Proteins which are not fully translated can either not be immobilized or not be probed because of their lack of either binding domain or immobilization-tag.

Chapter 3

Recent Developments in Single Molecule Force Spectroscopy

3.1 Summary

The atomic force microscope (AFM)-single molecule force spectroscopy (SMFS) community uses two approaches to probe molecules of interest. The classical approach consists of subcloning the protein of interest in-between repeats of IG- or other well-established fingerprint domains [125]. These large polyproteins are deposited on surfaces and adhere non-specifically to the cantilever. A non-functionalized cantilever moves over the surface and is picking up molecules by non-specific interactions. The fingerprint domains are necessary to identify a single interaction between cantilever and surface and to ensure that the protein of interest is fully loaded from its N- to C-terminus. This approach is usually very cumbersome and has a low yield in analyzable curves.
The alternative approach tries to combat this bottleneck and adapts techniques and advancements from fields like molecular biology and (bio)chemistry. It is possible to functionalize cantilever and surface site-specifically. This has the advantage of yielding a homogeneous surface with all molecules oriented in the same geometry. This strategy can be further improved by dividing the experimental setup of the protein of interest in two halfs. Instead of subcloning the protein of interest within a IG-scaffold, the protein is fused only to one binding domain of a receptor:ligand interaction. This fusion protein can be either anchored to the surface or the cantilever. The other part of the binding domain is coexpressed with a known fingerprint domain and immobilized to the opposite part of the AFM setup. By approaching the cantilever towards the surface the receptor:ligand interaction ensures a high number of specific tethers, compared to the non-specific approach. Retracting the cantilever ideally leads first to the unfolding of the fingerprint, then the protein of interest finally followed by complex rupture. With this approach it is possible to probe protein domains as well as receptor ligand interactions. Due to the advances in cloning it is possible to generate a fusion protein consisting either of fingerprint and binding domain of interest or a protein domain of interest with a characterized binding domain. Thus one can assemble the experimental construct like building bricks. The other advantage of this modular approach is the usability of the protein construct in other assays as well.
Associated publication P1 gives an overview of recent developments regarding AFM techniques employing site-specific pull-down strategies. Moreover, details about AFM theory, data handling and surface chemistry strategies are described.

3.2 Associated Publication P1

Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox

by

Wolfgang Ott*, Markus A. Jobst*, Constantin Schoeler*,

Hermann E. Gaub and Michael A. Nash

*(*contributed equally)*

published in

Journal of Structural Biology, 197 (1), 3-12, (2017)

Journal of Structural Biology 197 (2017) 3–12

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox

Wolfgang Ott [a,b,1], Markus A. Jobst [a,1], Constantin Schoeler [a,1], Hermann E. Gaub [a], Michael A. Nash [a,c,d,*]

[a] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, 81377 Munich, Germany
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[d] Department of Biosystems Science and Engineering, ETH-Zürich, 4058 Basel, Switzerland

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form 8 February 2016
Accepted 9 February 2016
Available online 9 February 2016

Keywords:
Atomic force microscopy
Single-molecule force spectroscopy
Molecular recognition
Biophysics

ABSTRACT

Single-molecule force spectroscopy sheds light onto the free energy landscapes governing protein folding and molecular recognition. Since only a single molecule or single molecular complex is probed at any given point in time, the technique is capable of identifying low-probability conformations within a large ensemble of possibilities. It furthermore allows choosing certain unbinding pathways through careful selection of the points at which the force acts on the protein or molecular complex. This review focuses on recent innovations in construct design, site-specific bioconjugation, measurement techniques, instrumental advances, and data analysis methods for improving workflow, throughput, and data yield of AFM-based single-molecule force spectroscopy experiments. Current trends that we highlight include customized fingerprint domains, peptide tags for site-specific covalent surface attachment, and polyproteins that are formed through mechanostable receptor–ligand interactions. Recent methods to improve measurement stability, signal-to-noise ratio, and force precision are presented, and theoretical considerations, analysis methods, and algorithms for analyzing large numbers of force–extension curves are further discussed. The various innovations identified here will serve as a starting point to researchers in the field looking for opportunities to push the limits of the technique further.

1. Introduction

The field began in earnest with the introduction of fluid cells for the (at that time) newly developed atomic force microscope (AFM) (Drake et al., 1989). The early 1990s then saw an explosion of the bio-AFM field, which opened the door to high-resolution imaging of proteins and cell surfaces under near-native conditions (Müller et al., 1995; Radmacher et al., 1996, 1992). Shortly thereafter came the realization that individual proteins and DNA molecules, or single receptor–ligand complexes, could be probed with the help of nano- to microscale force transducers (e.g., cantilevers, optically trapped beads, magnetically trapped beads) (Block et al., 1990; Florin et al., 1995; Lee et al., 1994a,b; Smith et al., 1992; Svoboda et al., 1993). It was furthermore discovered that natural polyproteins (e.g., Titin) with repetitive multi-domain structures provided regularly repeating saw-tooth like features in force extension data (Rief et al., 1997a). Artificial (i.e., recombinant) polyproteins quickly came into fashion as internal molecular controls for investigating mechanical properties of protein domains of interest. Since then, engineering of polyproteins has provided a wealth of information about mechanostable motifs in protein folds (Carrion-Vazquez et al., 1999; Oberhauser et al., 1998; Oesterhelt et al., 2000), directional dependence of protein mechanostability (Brockwell et al., 2003; Carrion-Vazquez et al., 2003; Dietz et al., 2006; Kim et al., 2011), and modulation of mechanostability by molecular recognition (Hu and Li, 2014).

Today, force spectroscopy and bio-AFM in general are well established as standard tools in the nanobiosciences, and are regularly used for investigating cell adhesion and cell surface properties (Helenius et al., 2008; Müller et al., 2009; Preiner et al., 2014; Tsukasaki et al., 2007; Wildling et al., 2012), interrogating membrane proteins (Beedle et al., 2015b; Janovjak et al., 2004; Müller, 2008; Müller and Engel, 2007), and measuring mechanical properties of proteins (Beedle et al., 2015a; Bu et al., 2012; Cao et al., 2011; del Rio et al., 2009; Geisler et al., 2010), polysaccharides (Kocun et al., 2011; Rief et al., 1997b) and DNA (Albrecht et al., 2003). Recent studies have already begun to characterize membrane proteins *in vivo* by probing their response to external forces on native living cells (Alsteens et al., 2010; Pfreundschuh et al.,

* Corresponding author.
E-mail address: michael.nash@lmu.de (M.A. Nash).
[1] Theses authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jsb.2016.02.011
1047-8477/

2015). There are a number of review articles that thoroughly cover the field from the early years (Carvalho et al., 2013; Casuso et al., 2011; Hoffmann and Dougan, 2012; Lee et al., 2007; Li and Cao, 2010; Marszalek and Dufrêne, 2012; Müller and Dufrêne, 2008; Neuman and Nagy, 2008; Noy, 2011; Rief and Grubmüller, 2002; Sirbuly et al., 2015; Woodside and Block, 2014).

Despite the high level of interest and well-developed method of AFM-SMFS (Single Molecule Force Spectroscopy), there have remained several limitations to the technique that prevent researchers from fully taking advantage of mechano-phenotyping of molecules and cell surfaces. Specifically, low experimental throughput and low yield of useable single-molecule interaction curves have both hampered the widespread adoption of the method, and its application for studying a large number of proteins. The purpose of this review is to highlight recent developments in bioconjugate chemistry, instrumentation, and data processing/algorithms which aim at improving the design process, yield, measurement quality and throughput of AFM-SMFS experiments.

2. Unfolding fingerprints

In typical AFM-SMFS experiments, many thousand force–extension curves are recorded, but only a fraction of these curves contain useable data that describe the behavior of a single molecule. Typically, the majority of curves (~80–99%) contain no interaction, a multiplicity of interactions that are difficult to interpret, or unspecific adhesion events as measurement artifacts. The experimenter is left searching for a needle in a haystack, looking for single-molecule interactions among a vast excess of unusable force–extension curves. In order to filter the data efficiently, the SMFS community has identified a broad range of proteins that can be used as specific identifiers in unfolding traces. We refer to these domains as 'fingerprints' because they provide a unique unfolding step or 'contour-length increment' of defined length that can be used as a filter during data processing. These fingerprint domains are typically globular protein domains with individual unfolding forces and length increments varying across a large range. This ability to choose the length increments and unfolding forces of the fingerprint domains has enabled the design of custom fusion proteins with well-controlled unfolding behaviors. Recent surveys of mechanical properties of different protein domains are provided by Sułkowska and Cieplak (2007), Hoffmann and Dougan (2012).

3. Receptor–ligand SMFS

Protein–protein and protein-small molecule interactions have been widely analyzed with SMFS. Reports of receptor–ligand SMFS include measurements on biotin–avidin (Florin et al., 1994; Lee et al., 1994a,b; Moy et al., 1994; Rico and Moy, 2007; Yuan et al., 2000), antigen–antibody interactions (Hinterdorfer et al., 1996; Morfill et al., 2007; Schwesinger et al., 2000) along with several other protein–protein or small molecule interactions (Lee et al., 2007; Mitchell et al., 2007; Schmidt et al., 2012).

One limitation in the standard method of receptor–ligand SMFS is that the signal lacks single-molecule specificity. Depending on the proteins involved and the experimental conditions (i.e., blocking/passivation steps), and since typically no fingerprint molecules are used, it can be difficult to differentiate non-specific interactions from specific protein-protein recognition. A second limitation of many receptor–ligand SMFS experiments is that pulling geometry is not strictly controlled. While in a standard polyprotein experiment, the force is applied strictly between the N- and C-termini of each domain, coupling of receptors and ligands to AFM tips and substrates is often done through amide linkages formed between amine groups on the proteins and activated NHS-ester groups on the surface or cantilever. This implicates a diversity of pulling geometries which are not strictly controlled, resulting in rupture force distributions that are smeared out or otherwise distorted.

4. Receptor–ligand SMFS with fingerprints

Our group has worked on improving the technique for receptor–ligand SMFS out of sheer necessity (Fig. 1). We were interested in studying a family of receptor–ligand proteins (i.e., cohesin–dockerin, Coh–Doc) involved in carbohydrate recognition and degradation by anaerobic bacteria (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012). These protein receptor–ligand complexes are responsible for building up large extracellular networks of structural scaffold proteins and enzymes. They are linked into these structural networks in well-defined and known orientations (e.g., N-terminal or C-terminal anchoring points). It is important to note that when pulling apart a receptor–ligand complex consisting of two proteins, there are four possible terminal pulling configurations (i.e., N-N′, N-C′, C-N′, C-C′) (Fig. 1B). Many of the Coh–Doc complexes we are interested in possess a clear 'physiological' pulling configuration found in nature, and 'non-physiological' or 'non-native' configurations. To understand their natural mechanical adaptations giving rise to their remarkable assembly strategy, we sought to characterize the mechanical stability of these receptor–ligand complexes in both their native and non-native loading configurations. We found a way to ensure specific interactions by basically combining two previously separate modes of AFM-SMFS (i.e., on polyproteins and receptor–ligand complexes). We fused the Coh and Doc domains separately to different fingerprint domains, and recombinantly produced each construct as a single fusion protein. The fingerprints serve two purposes: (1) they provide site-specific attachment sites through engineered cysteine residues or peptide ligation tags (see section 5) to strictly control loading geometry; (2) they provide predetermined increments in contour length which allows us to filter the datasets for specific single-molecule interactions (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2015, 2014; Stahl et al., 2012).

This configuration yields several advantages: We now have the ability to study mechanical stability of receptor–ligand pairs and unfolding of individual domains (i.e., the fingerprints) in a single-experiment with high yield and specificity, eliminating measurement artifacts. We also have a systematic and straightforward way to probe effects of pulling geometry on receptor–ligand unbinding, and to compare native and non-native pulling configurations. The gene design (i.e., N- or C-terminal fingerprint domains) directly reflects the conformation to be investigated. Furthermore, a specific protein domain of interest can now easily be fused to a mechanostable Coh–Doc receptor–ligand pair for characterization. Depending on the expected domain unfolding forces, an appropriately fitting protein receptor–ligand pair can be chosen from a wide range of well-characterized molecules (Table 1). We note that this table does not include every receptor–ligand probed by AFM. For an extensive list of receptor–ligands that were explored with AFM, see Lee et al. (2007). Currently, the mechanically most stable receptor–ligand pair is a Coh–Doc type III complex derived from *R. flavefaciens*, with loading-rate dependent rupture forces between 600 and 800 pN (Schoeler et al., 2015, 2014). Another interaction in a similar force range is the trimeric titin–telethonin complex described by Bertz et al. (2009).

5. Site-specific bioconjugation

Many polyprotein experiments rely on non-specific adsorption of polyproteins onto surfaces (e.g., mica, gold). Receptor–ligand

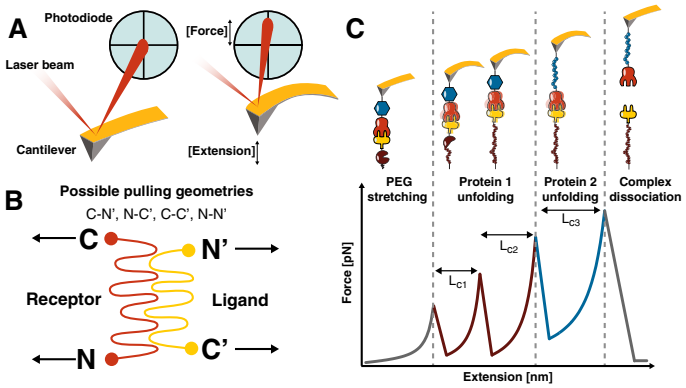


Fig. 1. Configuration for performing receptor–ligand SMFS with (poly)protein fingerprints. (A) Schematic of the measurement setup. The change of force is detected via the differential signal of the laser beam deflection on a quadrant photodiode. (B) For a protein complex consisting of two domains, 4 terminal pulling configurations are possible (N-N′, N-C′, C-N′, C-C′). (C) Fingerprints (brown and blue) are site-specifically and covalently attached to the cantilever and surface. Receptor (orange) and ligand (yellow) form a stable receptor–ligand complex. Note that the fingerprints can be individual sub-domains, or repetitive polyproteins in their own right. Shown is a typical force-extension trace with unfolding of the fingerprints, followed by rupture of the receptor ligand complex. In order to observe unfolding of the fingerprints in sufficient numbers, their most probable unfolding force should lie well below the most probable rupture force of the complex for the given loading rate.

Table 1
Overview of selected receptor–ligand pairs usable as specific handles for protein-based SMFS experiments. Rupture forces depend on immobilization sites for surface conjugation. Note that rupture forces can also vary depending on probe spring constants and loading rates. Abbreviations: NHS: N-hydroxysuccinimide; PEG: poly(ethylene glycol); Mal: maleimide; Cys: cysteine; CoA: coenzyme A; SFP: 4′-phosphopantetheinyl transferase; ybbR-Tag: peptide sequence DSLEFIASKLA; LF: low force unbinding path; HF: high force unbinding path. For the column 'immobilization method', the terminology X (Y) Z means: molecule X is attached to Z mediated by enzyme Y.

Protein handles Handle A:Handle B	Sizes (kDa)	Dissociation force (pN)	Immobilization method	Handle position (N/C)	References
Cohesin:dockerin I	15.4/8.3	122 ± 18.5	NHS-PEG5000-Mal/Cys	C:C	Stahl et al. (2012)
Cohesin:dockerin III	21.6/26.2	606 ± 54	NHS-PEG5000-Mal/Cys	N:C	Schoeler et al. (2015)
		111 ± 30 (LF)	NHS-PEG5000-Mal/CoA (SFP) ybbR	C:C	Schoeler et al. (2015)
		597 ± 67 (HF)	NHS-PEG5000-Mal/CoA (SFP) ybbR		
NiNTA:HIS6	0.2/0.8	153 ± 57	Gold-Cys	n.a.	Verbelen et al. (2007)
Avidin:biotin	66-69/0.2	160 ± 20	Biotinylated BSA	n.a.	Florin et al. (1994)
StrepTagII:streptavidin	1.1/52.8	253 ± 20	BSA/NHS-biotin	n.a.	Wong et al. (1999)
Streptavidin:biotin	52.8/0.2	200	Biotinylated BSA	n.a.	Rico and Moy (2007)
Calmodulin:CBP	16.7/1.1	16.5 ± 1.8	Pulldown via NI-NTA	n.a.	Junker and Rief (2009)
StrepTagII:mono-streptactin	1.1/58.4	116	NHS-PEG5000-Mal/Cys	C:C	Baumann et al. (2015)
		46	NHS-PEG5000-Mal/CoA (SFP) ybbR	N:C	
Anti-GCN4 sFv:GCN4(7P14P)	26.7/4.0	70	NHS-PEG5000-Mal/Cys	N:C	Morfill et al. (2007)
Anti-digoxigenin:digoxigenin	170/0.4	40	NHS-PEG6000	n.a.	Neuert et al. (2006)

AFM-SMFS, however, requires covalent immobilization of the two binding partners to the cantilever and surface, respectively, in order to avoid clogging of the molecules on the cantilever tip. Site-specific (i.e., residue specific) conjugation methods provide strict control over the pulling geometry and result in higher accuracy, precision and reproducibility, compared to conjugation methods resulting in a multiplicity of possible linkage sites (e.g., amine-targeting). Fig. 2 provides an overview of established surface chemistry strategies.

Another advantage of our modular system is the ability to use one construct (i.e., fingerprints with immobilization site) in all desired biochemical or biophysical assays, since immobilization relies on a PEG derivative, which is orthogonal to conventional specific pull down methods. It is compatible with a wide range of binding assays like Western Blotting, ITC, SPR, and ELISA.

The Ni-NTA:HIS_6-tag interaction can be used as force probe as well. This interaction has been employed as an adhesion sensor by probing a cell surface containing His-tagged protein. Since the His-tag is only located at one of the protein's termini, the insertion direction of the protein as well as it's position can be detected (Alsteens et al., 2013; Dupres et al., 2009; Pfreundschuh et al., 2015). This technique is especially useful since the His-tag can be used as a protein purification tag and simultaneously provides a single-molecule force handle.

5.1. Cysteines

Cysteines are relatively rare in proteins, making them attractive as a point mutation residue. The thiol side chain of cysteine is nucleophilic, and will spontaneously react with maleimide leaving groups at neutral pH. It can be used to site-specifically attach proteins to PEG coated surfaces for receptor–ligand AFM-SMFS. Alternatively, engineered cysteines can also be used as oligomerization sites to create disulfide-linked polyproteins, as was done for green fluorescent protein (GFP) (Dietz and Rief, 2006). However, cysteine/thiol-based protein conjugation has some drawbacks, including the tendency of cysteine-modified proteins to multimerize and ultimately aggregate over time, and incompatibility with proteins dis-

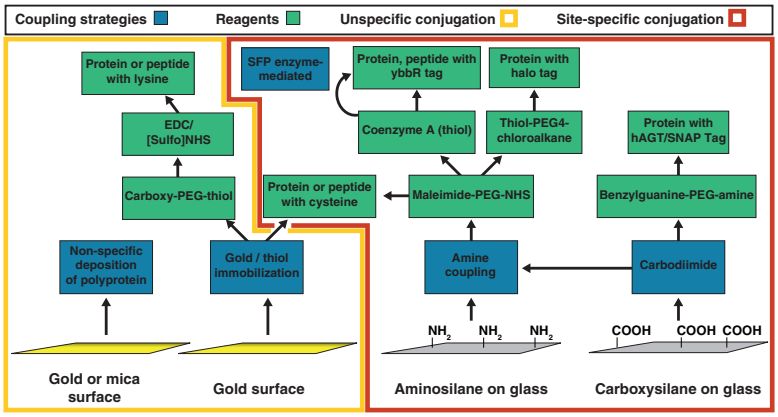


Fig. 2. Surface chemistry and bioconjugation strategies for single-molecule force spectroscopy. The diagram is by no means exhaustive and is roughly divided into site-specific conjugation methods that provide a single anchoring point for proteins to surfaces/cantilevers (right), and unspecific conjugation methods that provide a heterogeneity of loading configurations (i.e., a multiplicity of pulling points) (left).

playing cysteines on their surfaces in their wild-type form. Hence several other conjugation strategies were developed to overcome this challenge. Most of the newer techniques rely on N- or C-terminal attachment sites because the length of the requisite peptide tags or fusion domains makes inclusion into internal sites of a folded protein domain more challenging.

5.2. HaloTag

The active site of the haloalkane dehydrogenase (HaloTag) has been used to covalently immobilize proteins on chloroalkane surfaces. The unfolding forces of the HaloTag depend on its loading geometry (N-terminus: 131 pN; C-terminus: 491 pN). The domain provides an unfolding fingerprint of defined contour length, which also depends on the pulling geometry (N: 66 nm, C: 26.5 nm) (Popa et al., 2013).

5.3. hAGT/SNAP tag

The DNA repair protein O^6-alkylguanine–DNA-alkyltransferase (hAGT, SNAP-tag) binds benzylguanine covalently as a substrate, which can be attached to glass surfaces via an amino-polyethylene glycol (Kufer et al., 2005). With 22 kDa, the SNAP-tag is slightly smaller compared to the HaloTag (34 kDa).

5.4. SpyTag/Catcher

The versatile SpyTag/Catcher system can also be employed for site-specific surface immobilization. The linkage between SpyTag and Catcher is based on an internal protein interaction, which forms an isopeptide (covalent) bond. Based on this observation, the interaction was further developed and engineered, and now consists of a 13 amino acid large SpyTag and the binding domain Spy Catcher (Zakeri et al., 2012).

5.5. ybbR/SFP

The ybbR-Tag is an 11 amino acid protein sequence that is enzymatically linked to coenzyme A (CoA) by 4′-phosphopantetheinyl transferase (SFP) enzyme (Pippig et al., 2014; Yin et al., 2006; Yin et al., 2005). Both ybbR-Tag and the SpyTag/Catcher system have been shown to be N- and C-terminally active. Both tags can also be inserted internally, if the structure of the protein allows it, however, proper folding is not guaranteed and must be evaluated on a case-by-case basis.

5.6. Surface chemistry

Like the modular design of fingerprints and site-specific immobilization tags, surface chemistry can also be modularized to improve workflow. We note that the type of surface chemistry goes hand in hand with the design of the bioconjugation tags for protein production. Our standard approach follows the protocol described by Zimmermann et al. (2010): amino-silanized glass slides and cantilevers are functionalized with a hetero-bifunctional poly (ethylene glycol) (PEG) polymer with an N-hydroxysuccinimide group and a maleimide group at opposing ends. PEG coating provides a passivated surface that resists nonspecific protein adhesion, reducing background and artifacts during measurement. The entropic elasticity behavior of PEG (i.e., persistence length) is similar, although not equal to that of protein backbones, making it a suitable choice for surface conjugation in AFM-SMFS, without interfering too strongly with data interpretation. The maleimide group can then either be modified with CoA containing an inherent thiol group to proceed with ybbR/SFP chemistry, or alternatively directly be reacted with a protein domain displaying a reduced cysteine residue. The PEG incubation can be modified or extended depending on the requirements of the linker and the end group.

6. Advances in measurement techniques

Current advances in measurement resolution, instrument stability and accessible dynamic ranges open up new opportunities for measurements of biomolecules. Here we highlight recent innovations aimed at improving quality and precision of AFM-SMFS measurements.

6.1. Improved time resolution

In general, the timescales relevant for protein un-/folding and the corresponding timescale for thermally induced crossing of energy barriers are not fully detectable by common SMFS techniques, which typically resolve slower than 50 μs. Early on, the importance of developing high-speed AFM imaging and force spectroscopy through miniaturization of cantilevers with high resonance frequencies and low viscous drag coefficients was appreciated (Viani et al., 1999a,b). Nonetheless, only recent studies were able to overcome timescale limitations to observe, for example, extraordinarily slow protein misfolding transitions (~0.5 ms) using optical tweezers (Yu et al., 2015). Furthermore, advanced statistical methods extended optical tweezers SMFS time resolution to the ~10 μs range (Žoldák et al., 2013), and optimization of AFM cantilevers for SMFS has pushed the limit toward resolution on the microsecond timescale (Edwards et al., 2015). These developments allow experimentally accessible ranges to approach the lower limits of fast folding transition dynamics (Chung et al., 2012; Schuler and Hofmann, 2013), resolving short-lived intermediate states and yielding important insights into other fast conformational dynamics.

6.2. Bridging the timescale gap to steered molecular dynamics simulations

Recently, experimental measurements were brought into proximity (Dong and Sahin, 2011; He et al., 2012; Schoeler et al., 2015) or even overlap (Rico et al., 2013) with all atom steered molecular dynamics (SMD) simulations. Depending on the size and thus complexity of the simulated system, it has so far been possible to achieve SMD simulation timescales in the nanosecond to mid-microsecond range (Freddolino et al., 2008; Heymann and Grubmüller, 2001; Lee et al., 2009). Rico *et al.* developed a high speed force spectroscopy AFM based on an Ando-type high speed imaging AFM (Ando et al., 2001), with a high resonance frequency (600 kHz) miniature multilayer piezoelectric actuator (calibrated before each experiment and run in open loop mode), and a short cantilever with a high resonance frequency (550 kHz in liquid), and low viscous damping. This system was used to record protein unfolding data at extremely high speeds. To reduce hydrodynamic drag, the sample surface was tilted against the direction of the movement. With these improvements and data acquisition in the megahertz range, they were able to record meaningful and interpretable data at pulling speeds of up to 4000 μm/s, which is about 2–3 orders of magnitude faster than conventional methods and starts overlapping with the range of SMD simulations (Rico et al., 2013). Despite these successes, care must be taken because underdamped or 'ringing' cantilevers like the ones used here are not in agreement with the basic assumptions of the traditional SMFS framework, but can be improved by custom cantilever optimization procedures at the cost of time resolution (Edwards et al., 2015).

6.3. Long-term stability and force precision

Sophisticated measurements of complex biological systems or single molecules often require extraordinarily stable low-drift instruments, capable of continuous long-term data acquisition to gain sufficient and reliable statistics. Active stabilization techniques were developed to enable routine long-term stability and Ångstrom scale precision at room temperature for optical trap setups: differential sample position was measured and regulated with two independently stabilized and MHz modulated lasers, backscattered on sample and probe, and recorded separately on a single photodiode using lock-in amplifiers (Walder et al., 2015). This method is deemed applicable to surface-based and dual-beam optical traps, magnetic tweezers, AFM setups and optical microscopy, including super-resolution techniques.

AFM cantilever long-term stability and force precision can be increased even further by partially removing the reflective gold coating from the cantilever to dramatically reduce cantilever bending caused by the bimetallic effect (Churnside et al., 2012). Stability and precision improvements, which still retain high measurement bandwidths, enable and improve on picoscale force and sub-nanoscale motion measurements of molecular properties and dynamics in various biological systems. These may include groundbreaking investigations like the observation of single RNA polymerase base pair stepping (Abbondanzieri et al., 2005; Zhou et al., 2013), base pair unwinding of helicases (Cheng et al., 2011) and prion misfolding pathways (Yu et al., 2015, 2012). More details on long-term stability measurements and force precision are covered in the recent review of Edwards and Perkins (2016).

6.4. Mapping molecular recognition events: multiparametric imaging modes

The idea of mapping molecular recognition by simultaneously measuring surface topography and force–extension data ('force volume mapping' or 'affinity imaging') was introduced early (Hinterdorfer et al., 1996; Ludwig et al., 1997), and refined to remarkable temporal and spatial resolution. While these molecular recognition imaging techniques turned out to be a valuable tool for detecting and locating specific binding sites on surfaces, their development into dynamic recognition force imaging (Hinterdorfer and Dufrêne, 2006; Raab et al., 1999; Zhang et al., 2014) greatly increased temporal and spatial resolution, while still yielding information about surface elasticity and adhesion, as well as identifying biomolecules at the same time.

Multiparametric imaging modes can simultaneously detect physical properties of the surface and forces exerted on specific biomolecular binding sites. The AFM cantilever oscillates with amplitudes around 100 nm at sub- or low kilohertz frequencies to measure force–distance data, and simultaneously records image topography and other surface properties at sub- or low hertz line-scanning frequencies. The recorded force and topography data is collected orders of magnitude faster compared to force volume mapping methods, yielding imaging speeds comparable to conventional AFM imaging methods (Alsteens et al., 2012; Pfreundschuh et al., 2014). Another benefit of this method is that a large range of loading rates for receptor–ligand dissociation events can be probed in a single experiment, due to the largely varying cantilever tip velocities. Recently, this method was applied to gain nm-scale resolution imaging data of a G protein-coupled receptor (PAR1) in proteoliposomes while characterizing their ligand-binding energy landscape (Alsteens et al., 2015) from loading rates ranging between 1e3 and 1e6 pN/s, already two orders of magnitude higher than conventional force–distance based SMFS. Another recent study demonstrates the ability of this technique to distinguish two different binding events on opposite sides of engineered PAR1 by their unbinding force, and thereby determine their orientation within the lipid bilayer (Pfreundschuh et al., 2015).

6.5. Lateral force sensors

A slightly different approach developed a T-shaped cantilever (Dong et al., 2009; Dong and Sahin, 2011) to drive it at its flexural resonance frequency (~9 kHz) and record force data from cantilever torsion, resulting in a lateral laser deflection signal that was acquired while imaging the sample in conventional tapping mode. Due to the cantilever's high torsional resonance (~115 kHz), unbinding dynamics could be measured at the

microsecond timescale and at extraordinarily high loading rates of up to nearly 1e9 pN/s (Dong and Sahin, 2011), about four orders of magnitude faster than conventional SMFS. Force curves and therefore unbinding events and their corresponding force values could be mapped with high spatial and temporal resolution, while providing AFM images that were simultaneously recorded as surface topography. Mechanical elasticity properties of the substrate were also detected in the phase signal.

7. Theory and data analysis

7.1. The data analysis problem

Technical advances greatly increasing the throughput of AFM-SMFS measurements have made automated data analysis protocols an essential requirement. In practice, researchers face the problem of extracting meaningful single molecule signal from large datasets that contain an abundance of unusable data. The use of well-defined fingerprint domains with known unfolding patterns facilitates this procedure greatly. To avoid tedious and time-consuming manual sorting of thousands of data traces, and potential introduction of bias into the data analysis procedure, algorithms which identify the fingerprint unfolding length increments and classify the data correspondingly have been developed and implemented with success (Bosshart et al., 2012; Jobst et al., 2015; Kuhn et al., 2005; Puchner et al., 2008).

7.2. Polymer elasticity models and contour length transformations

Single molecule force measurements generally only gain access to a protein's extension under a given force. The stochastic nature of domain unfolding or complex dissociation under force as well as the non-linear elastic behavior of the polymer backbone chain makes analysis in force-extension space difficult. The same unfolding event is observed over a range of different positions in force–extension curves for multiple measurement cycles as shown in Fig. 3B i.

From a physicist's point of view, mechanical stretching of an unfolded protein domain is described by polymer elasticity models such as the worm-like chain (WLC) (Bustamante et al., 1994), the freely jointed chain (FJC) (Ortiz and Hadziioannou, 1999), or the freely rotating chain (FRC) model (Livadaru et al., 2003). These models contain the free contour length L of the polymer, including surface tethers and unfolded protein backbone, as a parameter. The free contour length is simply the length of the polypeptide along the contour of the biopolymer chain, given a specific folding state (e.g., Fig. 3A). Under a set of physically relevant constraints (L, x, $F > 0$, $x < L$), these elasticity models provide one-to-one mappings from force–extension space into force-contour length space. The models can be solved for the contour length parameter (Jobst et al., 2013; Puchner et al., 2008), yielding an expression for the contour length as a function of force and extension $L(F,x)$. This function can be used to transform force–extension traces from constant speed or force clamp/ramp experiments into contour length space (Fig. 3B ii). The calculated contour length then can be binned (Fig. 3B iii), aligned, and subsequently averaged to precisely locate energy barriers (Fig. 3B iv) along a protein's unfolding pathway, and to classify data sets based on unfolding patterns. This idea was first proposed by Puchner et al. (2008) and has been successfully applied in multiple AFM-SMFS studies (Jobst et al., 2015, 2013; Otten et al., 2014; Schoeler et al., 2014; Stahl et al., 2012; Thoma et al., 2015).

7.3. Worm-like chains, freely-rotating chains and beyond

The WLC model accurately describes a protein's stretching response for forces up to approximately 150 pN. While many protein unfolding or dissociation events take place well within this force regime, some interactions like titin Ig domain unfolding (Rief et al., 1997a), cohesin unfolding (Valbuena et al., 2009), disso-

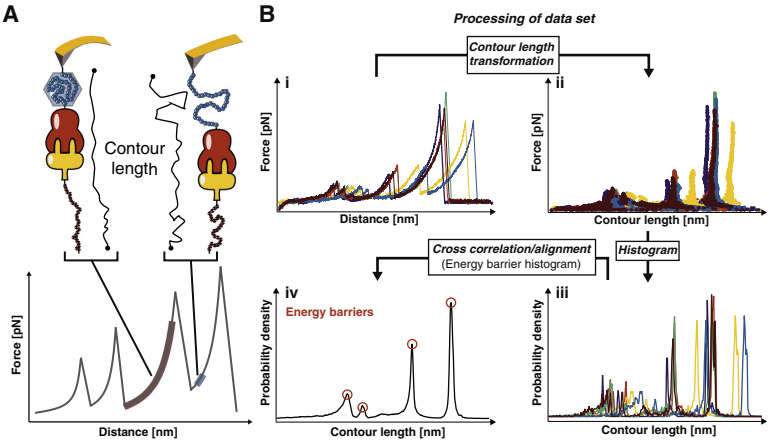


Fig. 3. Assembly of contour length histograms for screening AFM-SMFS datasets. (A) Force–extension traces are transformed into contour length space using an appropriate polymer elasticity model. The choice of the model depends on the force range. (B) Following transformation, the data (i) are plotted in force-contour length space (ii). Force and contour length thresholds are applied and the data are histogrammed (projected onto contour length axis) with an appropriate bin width, i.e., nanometer scale, to obtain the diagram in (iii). Each trace analyzed this way can be searched for a specific contour length increment (distance between two peaks in the probability density vs. contour length plot) corresponding to one of the fingerprints. To obtain a master histogram describing all the observed increments in a dataset, individual histograms reflecting a specific unfolding pathway are aligned by cross-correlation and offsetting along the contour-length axis (iv).

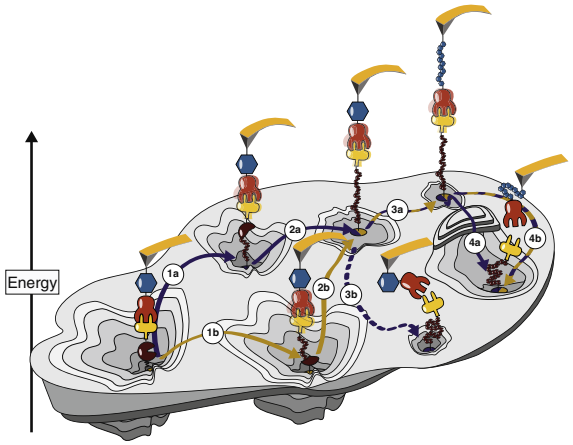


Fig. 4. Schematic depiction of an (un)folding energy landscape. The bound state of a protein receptor–ligand complex can be thought of as a Brownian particle confined to a complex multidimensional energy landscape. At equilibrium, the system can escape the bound state driven by thermal fluctuations. This escape can occur along any pathway on the energy landscape. When measuring the thermal off-rate with bulk assays such as surface plasmon resonance biosensors, a weighted average of all thermally accessible pathways is obtained. In a single-molecule pulling experiment, however, a small subset of pathways is selected, which is defined by the projection of the energy landscape onto the pulling coordinate as illustrated by paths 1–3. Caution is required when comparing data obtained from single molecule techniques with bulk data. In cases where SMFS probes a steep pathway with a high free energy barrier, the fitted zero-force off rate may greatly differ from values obtained by bulk techniques. Path 4 illustrates the thermal escape (4b) versus the forced pathway across an additional energy barrier (4a) by the AFM cantilever.

ciation of skeletal muscle titin–telethonin bonds (Bertz et al., 2009) or dissociation of cellulosomal adhesion complexes (Schoeler et al., 2015, 2014) exhibit much higher unfolding or rupture forces. To adequately describe the elastic response of polymers in such high force regimes, models beyond the standard WLC are required. To address this shortcoming, Hugel et al. (2005) developed quantum mechanical corrections for polymer elasticity models to account for polypeptide backbone stretching at high forces. These corrections can be applied to obtain the contour length at zero force L_0 (Puchner et al., 2008).

Livadaru et al. proposed a more sophisticated model exhibiting three distinct regimes for a protein's stretching response as a function of the applied force (Livadaru et al., 2003). For AFM based SMFS, however, mainly the medium to high force regimes are relevant. The medium force regime of protein stretching, roughly between 10 and 125 pN, exhibits classical WLC stretching behavior, whereas the high force regime shows the behavior of a discrete chain, where the stretching response is independent of the persistence length. This model is most suitable for studying high force interactions, especially when combined with the aforementioned quantum mechanical corrections for backbone stretching.

8. Kinetic and energetic parameters

In dynamic force spectroscopy of receptor–ligand pairs, kinetic and energetic parameters of the complex are of interest. The method most prominently used to extract this information from SMFS experiments is to vary the loading rate by measuring the rupture forces at different pulling speeds in constant speed mode (Baumann et al., 2015; Schoeler et al., 2014; Stahl et al., 2012), or with different slopes in force ramp mode (Oberhauser et al., 2001). The obtained rupture force data are then assembled into a dynamic force spectrum, a plot of most probable rupture forces against their corresponding loading rates. In their comprehensive guide to analysis of SMFS data sets, Noy and Friddle (2013) explain the basic physics of bond stretching. An SMFS measurement corresponds to the stretching of multiple elastic components in series, including the projection of the bond potential onto the pulling axis, the cantilever modeled as a harmonic spring and potential linker molecules with nonlinear elasticity deviating from those under investigation. Such a scenario gives rise to bound and unbound states separated by free energy barriers. By pulling on the harmonic spring, this energy landscape is constantly modulated. Since thermal fluctuations are orders of magnitude faster than changes in the external force, the transition from a bound to an unbound state is thermally driven in common loading rate regimes, as described by Bell (1978), Evans and Ritchie (1997), Izrailev et al. (1997). These models describe a linear dependence of the rupture force on the natural logarithm of the loading rate and give access to the zero force off rate k_0 (exponentially amplified under force) and the distance to the transition state Δx. Theoreticians extended this framework and accounted for modulation of Δx by the applied force (Dudko et al., 2006), and the possibility of rebinding at slow loading rates (Friddle et al., 2012). These newer models predict a nonlinear dependence of the most probable rupture force on the loading rate and give the height of the free energy barrier to unbinding ΔG as an additional parameter. Such non-linear trends were observed experimentally, and a comprehensive list of such data sets is given in Friddle et al. (2012). Joint experimental and computational data sets were also analyzed in recent studies (Rico et al., 2013; Schoeler et al., 2015). As Noy and Friddle (2013) point out, these models should only be used if the force spectrum of interest indeed exhibits a non-linear trend. If this is not given, fitting non-linear models results in non-meaningful fit parameters and the phenomenological model should be used instead.

Although in both bulk measurements and single molecule force measurements at common loading rates, the unbinding process is

thermally driven, caution is required when comparing their data. While at unbiased equilibrium, all thermally accessible pathways from the bound state are sampled and the off rate is consequently measured as a weighted average, single molecule force measurements select only a small subset of these pathways due to the defined pulling geometry, as illustrated by paths 1–3 in Fig. 4. In cases where the energy landscape is highly asymmetric and the pulling experiment probes a steep pathway, the off rates obtained from single molecule *vs.* bulk measurements might differ greatly (see Fig. 4, paths 4a vs 4b).

9. Summary and outlook

We highlighted recent advances in experimental design, molecular design, sample preparation, measurement and analysis methods for AFM-SMFS on polyproteins and receptor–ligand complexes. We summarized site-specific bioconjugation strategies to obtain well-defined pulling geometries for improved reliability and reproducibility of experiments. We also highlighted receptor–ligand pairs with high mechanical strength (e.g., cohesin–dockerin), and their application as specific pulling handles in AFM-SMFS for improving experimental throughput and curve yield. Finally, we touched on recent innovations in positional control and cantilever microfabrication for improving time and force resolution and stability of the measurement, on emerging techniques for mapping force responses of surfaces to their topologies, and we discussed theoretical considerations for analyzing large numbers of curves.

In the future, there remain several technical challenges that need to be addressed. One of the limitations of AFM is that it covers a relatively high force range, yet there exist a multitude of biological interactions in the low-force regime that are of interest. Further technical advances in instrument design, cantilever fabrication, and feedback control might further improve force resolution and thereby enable such experiments. A second area for improvement involves sample throughput and parallel screening. With the development of more elaborate, sophisticated and well defined surface immobilization strategies and protein handles, significant gains in throughput can be envisioned. Innovations of the chemistry in combination with efficient data analysis protocols and state of the art instrumentation may pave the way towards in depth study of complex, multi-domain protein systems.

These advances in experimental design and throughput would greatly benefit from refined theoretical frameworks that account for parameters such as cantilever stiffness and ringing whilst maintaining analytical tractability. Consequently, with improved methodology we anticipate the community will be able to address an even wider range of questions about mechanical adaptations of proteins and protein complexes in the future.

Acknowledgments

The authors gratefully acknowledge funding from an advanced Grant to HEG from the European Research Council (Cellufuel Grant 294438), and from a Society in Science – Branco Weiss Fellowship administered by ETH Zürich, Switzerland to MAN. We are also thankful for the financial support of the Deutsche Forschungsgemeinschaft SFB 863.

References

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., Block, S.M., 2005. Direct observation of base-pair stepping by RNA polymerase. Nature 438, 460–465. http://dx.doi.org/10.1038/nature04268.

Albrecht, C., Blank, K., Lalic-Multhaler, M., Hirler, S., Mai, T., Gilbert, I., Schiffmann, S., Bayer, T., Clausen-Schaumann, H., Gaub, H.E., 2003. DNA: a programmable force sensor. Science 301, 367–370. http://dx.doi.org/10.1126/science.1084713.

Alsteens, D., Garcia, M.C., Lipke, P.N., Dufrêne, Y.F., 2010. Force-induced formation and propagation of adhesion nanodomains in living fungal cells. Proc. Natl. Acad. Sci. USA 107, 20744–20749. http://dx.doi.org/10.1073/pnas.1013893107.

Alsteens, D., Dupres, V., Yunus, S., Latgé, J.-P., Heinisch, J.J., Dufrêne, Y.F., 2012. High-resolution imaging of chemical and biological sites on living cells using peak force tapping atomic force microscopy. Langmuir 28, 16738–16744. http://dx.doi.org/10.1021/la303891j.

Alsteens, D., Trabelsi, H., Soumillion, P., Dufrêne, Y.F., 2013. Multiparametric atomic force microscopy imaging of single bacteriophages extruding from living bacteria. Nat. Commun. 4, 1–7. http://dx.doi.org/10.1038/ncomms3926.

Alsteens, D., Pfreundschuh, M., Zhang, C., Spoerri, P.M., Coughlin, S.R., Kobilka, B.K., Müller, D.J., 2015. Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape. Nat. Methods 12, 845–851. http://dx.doi.org/10.1038/nmeth.3479.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K., Toda, A., 2001. A high-speed atomic force microscope for studying biological macromolecules. Proc. Natl. Acad. Sci. USA 98, 12468–12472. http://dx.doi.org/10.1073/pnas.211400898.

Baumann, F., Bauer, M.S., Milles, L.F., Alexandrovich, A., Gaub, H.E., Pippig, D.A., 2015. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. Nat. Nanotechnol. 1–7. http://dx.doi.org/10.1038/nnano.2015.231.

Beedle, A.E.M., Lezamiz, A., Stirnemann, G., Garcia-Manyes, S., 2015a. The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. Nat. Commun. 6, 1–9. http://dx.doi.org/10.1038/ncomms8894.

Beedle, A.E.M., Williams, A., Relat-Goberna, J., Garcia-Manyes, S., 2015b. Mechanobiology – chemical origin of membrane mechanical resistance and force-dependent signaling. Curr. Opin. Chem. Biol. 29, 87–93. http://dx.doi.org/10.1016/j.cbpa.2015.09.019.

Bell, G.I., 1978. Models for the specific adhesion of cells to cells. Science 200, 618–627. http://dx.doi.org/10.1126/science.347575.

Bertz, M., Wilmanns, M., Rief, M., 2009. The titin–telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. Proc. Natl. Acad. Sci. USA 106, 13307–13310. http://dx.doi.org/10.1073/pnas.0902312106.

Block, S.M., Goldstein, L., Schnapp, B.J., 1990. Bead movement by single kinesin molecules studied with optical tweezers. Nature 348, 348–352. http://dx.doi.org/10.1038/348348a0.

Bosshart, P.D., Frederix, P.L.T.M., Engel, A., 2012. Reference-free alignment and sorting of single-molecule force spectroscopy data. Biophys. J. 102, 2202–2211. http://dx.doi.org/10.1016/j.bpj.2012.03.027.

Brockwell, D.J., Paci, E., Zinober, R.C., Beddard, G.S., Olmsted, P.D., Smith, D.A., Perham, R.N., Radford, S.E., 2003. Pulling geometry defines the mechanical resistance of a beta-sheet protein. Nat. Struct. Biol. 10, 731–737. http://dx.doi.org/10.1038/nsb968.

Bu, T., Wang, H.-C.E., Li, H., 2012. Single molecule force spectroscopy reveals critical roles of hydrophobic core packing in determining the mechanical stability of protein GB1. Langmuir 28, 12319–12325. http://dx.doi.org/10.1021/la301940g.

Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S., 1994. Entropic elasticity of lambda-phage DNA. Science 265, 1599–1600. http://dx.doi.org/10.1126/science.8079175.

Cao, Y., Li, Y.D., Li, H., 2011. Enhancing the mechanical stability of proteins through a cocktail approach. Biophys. J. 100, 1794–1799. http://dx.doi.org/10.1016/j.bpj.2011.02.030.

Carrion-Vazquez, M., Oberhauser, A.F., Fowler, S.B., Marszalek, P.E., Broedel, S.E., Clarke, J., Fernandez, J.M., 1999. Mechanical and chemical unfolding of a single protein: a comparison. Proc. Natl. Acad. Sci. USA 96, 3694–3699. http://dx.doi.org/10.1073/pnas.96.7.3694.

Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P.E., Oberhauser, A.F., Fernandez, J.M., 2003. The mechanical stability of ubiquitin is linkage dependent. Nat. Struct. Biol. 10, 738–743. http://dx.doi.org/10.1038/nsb965.

Carvalho, F.A., Martins, I.C., Santos, N.C., 2013. Atomic force microscopy and force spectroscopy on the assessment of protein folding and functionality. Arch. Biochem. Biophys. 531, 116–127. http://dx.doi.org/10.1016/j.abb.2012.11.007.

Casuso, I., Rico, F., Scheuring, S., 2011. Biological AFM: where we come from – where we are – where we may go. J. Mol. Recogn. 24, 406–413. http://dx.doi.org/10.1002/jmr.1081.

Cheng, W., Arunajadai, S.G., Moffitt, J.R., Tinoco, I.J., Bustamante, C., 2011. Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. Science 333, 1746–1749. http://dx.doi.org/10.1126/science.1206023.

Chung, H.S., McHale, K., Louis, J.M., Eaton, W.A., 2012. Single-molecule fluorescence experiments determine protein folding transition path times. Science 335, 981–984. http://dx.doi.org/10.1126/science.1215768.

Churnside, A.B., Sullan, R.M.A., Nguyen, D.M., Case, S.O., Bull, M.S., King, G.M., Perkins, T.T., 2012. Routine and timely sub-picoNewton force stability and precision for biological applications of atomic force microscopy. Nano Lett. 12, 3557–3561. http://dx.doi.org/10.1021/nl301166w.

del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., Sheetz, M.P., 2009. Stretching single Talin rod molecules activates vinculin binding. Science 323, 638–641. http://dx.doi.org/10.1126/science.1162912.

Dietz, H., Rief, M., 2006. Protein structure by mechanical triangulation. Proc. Natl. Acad. Sci. USA 103, 1244–1247. http://dx.doi.org/10.1073/pnas.0509217103.

Dietz, H., Berkemeier, F., Bertz, M., Rief, M., 2006. Anisotropic deformation response of single protein molecules. Proc. Natl. Acad. Sci. USA 103, 12724–12728. http://dx.doi.org/10.1073/pnas.0602995103.

Dong, M., Sahin, O., 2011. A nanomechanical interface to rapid single-molecule interactions. Nat. Commun. 2, 1–6. http://dx.doi.org/10.1038/ncomms1246.

Dong, M., Husale, S., Sahin, O., 2009. Determination of protein structural flexibility by microsecond force spectroscopy. Nat. Nanotechnol. 4, 514–517. http://dx.doi.org/10.1038/nnano.2009.156.

Drake, B., Prater, C.B., Weisenhorn, A.L., Gould, S.A.C., Albrecht, T.R., Quate, C.F., Cannell, D.S., Hansma, H.G., Hansma, P.K., 1989. Imaging crystals, polymers, and processes in water with the atomic force microscope. Science 243, 1386–1389. http://dx.doi.org/10.1126/science.2928794.

Dudko, O.K., Hummer, G., Szabo, A., 2006. Intrinsic rates and activation free energies from single-molecule pulling experiments. Phys. Rev. Lett. 96, 108101. http://dx.doi.org/10.1103/PhysRevLett. 96.108101.

Dupres, V., Alsteens, D., Wilk, S., Hansen, B., Heinisch, J.J., Dufrêne, Y.F., 2009. The yeast Wsc1 cell surface sensor behaves like a nanospring in vivo. Nat. Chem. Biol. 5, 857–862. http://dx.doi.org/10.1038/nchembio.220.

Edwards, D.T., Perkins, T.T., 2016. Optimizing force spectroscopy by modifying commercial cantilevers: improved stability, precision, and temporal resolution. J. Struct. Biol. 1–13. http://dx.doi.org/10.1016/j.jsb.2016.01.009.

Edwards, D.T., Faulk, J.K., Sanders, A.W., Bull, M.S., Walder, R., LeBlanc, M.-A., Sousa, M.C., Perkins, T.T., 2015. Optimizing 1-μs-resolution single-molecule force spectroscopy on a commercial atomic force microscope. Nano Lett. 15, 7091–7098. http://dx.doi.org/10.1021/acs.nanolett.5b03166.

Evans, E., Ritchie, K., 1997. Dynamic strength of molecular adhesion bonds. Biophys. J. 72, 1541–1555. http://dx.doi.org/10.1016/S0006-3495(97)78802-7.

Florin, E.L., Moy, V.T., Gaub, H.E., 1994. Adhesion forces between individual ligand–receptor pairs. Science 264, 415–417. http://dx.doi.org/10.1126/science.8153628.

Florin, E.L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V.T., Gaub, H.E., 1995. Sensing specific molecular-interactions with the atomic-force microscope. Biosens. Bioelectron. 10, 895–901. http://dx.doi.org/10.1016/0956-5663(95)99227-C.

Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K., 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. 94, L75–L77. http://dx.doi.org/10.1529/biophysj.108.131565.

Friddle, R.W., Noy, A., De Yoreo, J.J., 2012. Interpreting the widespread nonlinear force spectra of intermolecular bonds. Proc. Natl. Acad. Sci. USA 109, 13573–13578. http://dx.doi.org/10.1073/pnas.1202946109.

Geisler, M., Xiao, S., Puchner, E.M., Gräter, F., Hugel, T., 2010. Controlling the structure of proteins at surfaces. J. Am. Chem. Soc. 132, 17277–17281. http://dx.doi.org/10.1021/ja107212z.

He, C., Genchev, G.Z., Lu, H., Li, H., 2012. Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. J. Am. Chem. Soc. 134, 10428–10435. http://dx.doi.org/10.1021/ja3003205.

Helenius, J., Heisenberg, C.-P., Gaub, H.E., Müller, D.J., 2008. Single-cell force spectroscopy. J. Cell Sci. 121, 1785–1791. http://dx.doi.org/10.1242/jcs.030999.

Heymann, B., Grubmüller, H., 2001. Molecular dynamics force probe simulations of antibody/antigen unbinding: entropic control and nonadditivity of unbinding forces. Biophys. J. 81, 1295–1313. http://dx.doi.org/10.1016/S0006-3495(01)75787-6.

Hinterdorfer, P., Dufrêne, Y.F., 2006. Detection and localization of single molecular recognition events using atomic force microscopy. Nat. Methods 3, 347–355. http://dx.doi.org/10.1038/nmeth871.

Hinterdorfer, P., Baumgartner, W., Gruber, H.J., Schilcher, K., Schindler, H., 1996. Detection and localization of individual antibody–antigen recognition events by atomic force microscopy. Proc. Natl. Acad. Sci. USA 93, 3477–3481. http://dx.doi.org/10.1073/pnas.93.8.3477.

Hoffmann, T., Dougan, L., 2012. Single molecule force spectroscopy using polyproteins. Chem. Soc. Rev. 41, 4781–4796. http://dx.doi.org/10.1039/c2cs35033e.

Hu, X., Li, H., 2014. Force spectroscopy studies on protein–ligand interactions: a single protein mechanics perspective. FEBS Lett. 588, 3613–3620. http://dx.doi.org/10.1016/j.febslet.2014.04.009.

Hugel, T., Rief, M., Seitz, M., Gaub, H.E., Netz, R.R., 2005. Highly stretched single polymers: atomic-force-microscope experiments versus ab-initio theory. Phys. Rev. Lett. http://dx.doi.org/10.1103/PhysRevLett.94.048301.

Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., Schulten, K., 1997. Molecular dynamics study of unbinding of the avidin–biotin complex. Biophys. J. 72, 1568–1581. http://dx.doi.org/10.1016/S0006-3495(97)78804-0.

Janovjak, H., Struckmeier, J., Hubain, M., Kedrov, A., Kessler, M., Müller, D.J., 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure 12, 871–879. http://dx.doi.org/10.1016/j.str.2004.03.016.

Jobst, M.A., Schoeler, C., Malinowska, K., Nash, M.A., 2013. Investigating receptor–ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. J. Vis. Exp., e50950 http://dx.doi.org/10.3791/50950.

Jobst, M.A., Milles, L.F., Schoeler, C., Ott, W., Fried, D.B., Bayer, E.A., Gaub, H.E., Nash, M.A., 2015. Resolving dual binding conformations of cellulosome cohesin–dockerin complexes using single-molecule force spectroscopy. Elife 4, 1031. http://dx.doi.org/10.7554/eLife.9.

Junker, J.P., Rief, M., 2009. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. Proc. Natl. Acad. Sci. USA 106, 14361–14366. http://dx.doi.org/10.1073/pnas.0904654106.

Kim, M., Wang, C.-C., Benedetti, F., Rabbi, M., Bennett, V., Marszalek, P.E., 2011. Nanomechanics of streptavidin hubs for molecular materials. Adv. Mater. 23, 5684–5688. http://dx.doi.org/10.1002/adma.201103316.

Kocun, M., Grandbois, M., Cuccia, L.A., 2011. Single molecule atomic force microscopy and force spectroscopy of chitosan. Colloids Surf. B Biointerfaces 82, 470–476. http://dx.doi.org/10.1016/j.colsurfb.2010.10.004.

Kufer, S.K., Dietz, H., Albrecht, C., Blank, K., Kardinal, A., Rief, M., Gaub, H.E., 2005. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. Eur. Biophys. J. 35, 72–78. http://dx.doi.org/10.1007/s00249-005-0010-1.

Kuhn, M., Janovjak, H., Hubain, M., Müller, D.J., 2005. Automated alignment and pattern recognition of single-molecule force spectroscopy data. J. Microscopy 218, 125–132. http://dx.doi.org/10.1111/j.1365-2818.2005.01478.x.

Lee, C.-K., Chrisey, L.A., Colton, R.J., 1994a. Direct measurement of the forces between complementary strands of DNA. Science. http://dx.doi.org/10.1126/science.7973628.

Lee, G.U., Kidwell, D.A., Colton, R.J., 1994b. Sensing discrete streptavidin–biotin interactions with atomic force microscopy. Langmuir. http://dx.doi.org/10.1021/la00014a003.

Lee, C.-K., Wang, Y.-M., Huang, L.-S., Lin, S., 2007. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein-ligand interaction. Micron 38, 446–461. http://dx.doi.org/10.1016/j.micron.2006.06.014.

Lee, E.H., Hsin, J., Sotomayor, M., Comellas, G., Schulten, K., 2009. Discovery through the computational microscope. Structure 17, 1295–1306. http://dx.doi.org/10.1016/j.str.2009.09.001.

Li, H., Cao, Y., 2010. Protein mechanics: from single molecules to functional biomaterials. Acc. Chem. Res. 43, 1331–1341. http://dx.doi.org/10.1021/ar100057a.

Livadaru, L., Netz, R.R., Kreuzer, H.J., 2003. Stretching response of discrete semiflexible polymers. Macromolecules 36, 3732–3744. http://dx.doi.org/10.1021/ma020751g.

Ludwig, M., Dettmann, W., Gaub, H.E., 1997. Atomic force microscope imaging contrast based on molecular recognition. Biophys. J. 72, 445–448. http://dx.doi.org/10.1016/S0006-3495(97)78685-5.

Marszalek, P.E., Dufrêne, Y.F., 2012. Stretching single polysaccharides and proteins using atomic force microscopy. Chem. Soc. Rev. 41, 3523–3534. http://dx.doi.org/10.1039/c2cs15329g.

Mitchell, G., Lamontagne, C.-A., Lebel, R., Grandbois, M., Malouin, F., 2007. Single-molecule dynamic force spectroscopy of the fibronectin–heparin interaction. Biochem. Biophys. Res. Commun. 364, 595–600. http://dx.doi.org/10.1016/j.bbrc.2007.10.034.

Morfill, J., Blank, K., Zahnd, C., Luginbühl, B., Kühner, F., Gottschalk, K.-E., Plückthun, A., Gaub, H.E., 2007. Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. Biophys. J. 93, 3583–3590. http://dx.doi.org/10.1529/biophysj.107.112532.

Moy, V.T., Florin, E.L., Gaub, H.E., 1994. Intermolecular forces and energies between ligands and receptors. Science 266, 257–259. http://dx.doi.org/10.1126/science.7939660.

Müller, D.J., 2008. AFM: a nanotool in membrane biology. Biochemistry 47, 7986–7998. http://dx.doi.org/10.1021/bi800753x.

Müller, D.J., Dufrêne, Y.F., 2008. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. Nat. Nanotechnol. 3, 261–269. http://dx.doi.org/10.1038/nnano.2008.100.

Müller, D.J., Engel, A., 2007. Atomic force microscopy and spectroscopy of native membrane proteins. Nat. Protocols 2, 2191–2197. http://dx.doi.org/10.1038/nprot.2007.309.

Müller, D.J., Büldt, G., Engel, A., 1995. Force-induced conformational change of bacteriorhodopsin. J. Mol. Biol. 249, 239–243. http://dx.doi.org/10.1006/jmbi.1995.0292.

Müller, D.J., Helenius, J., Alsteens, D., Dufrêne, Y.F., 2009. Force probing surfaces of living cells to molecular resolution. Nat. Chem. Biol. 5, 383–390. http://dx.doi.org/10.1038/nchembio.181.

Neuert, G., Albrecht, C., Pamir, E., Gaub, H.E., 2006. Dynamic force spectroscopy of the digoxigenin–antibody complex. FEBS Lett. 580, 505–509. http://dx.doi.org/10.1016/j.febslet.2005.12.052.

Neuman, K.C., Nagy, A., 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat. Methods 5, 491–505. http://dx.doi.org/10.1038/nmeth.1218.

Noy, A., 2011. Force spectroscopy 101: how to design, perform, and analyze an AFM-based single molecule force spectroscopy experiment. Curr. Opin. Chem. Biol. 15, 710–718. http://dx.doi.org/10.1016/j.cbpa.2011.07.020.

Noy, A., Friddle, R.W., 2013. Practical single molecule force spectroscopy: how to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope. Methods 60, 142–150. http://dx.doi.org/10.1016/j.ymeth.2013.03.014.

Oberhauser, A.F., Marszalek, P.E., Erickson, H.P., Fernandez, J.M., 1998. The molecular elasticity of the extracellular matrix protein tenascin. Nature 393, 181–185. http://dx.doi.org/10.1038/30270.

Oberhauser, A.F., Hansma, P.K., Carrion-Vazquez, M., Fernandez, J.M., 2001. Stepwise unfolding of titin under force-clamp atomic force microscopy. Proc. Natl. Acad. Sci. USA 98, 468–472. http://dx.doi.org/10.1073/pnas.021321798.

Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E., Müller, D.J., 2000. Unfolding pathways of individual bacteriorhodopsins. Science 288, 143–146. http://dx.doi.org/10.1126/science.288.5463.143.

Ortiz, C., Hadziioannou, G., 1999. Entropic elasticity of single polymer chains of poly (methacrylic acid) measured by atomic force microscopy. Macromolecules 32, 780–787. http://dx.doi.org/10.1021/ma981245n.

Otten, M., Ott, W., Jobst, M.A., Milles, L.F., Verdorfer, T., Pippig, D.A., Nash, M.A., Gaub, H.E., 2014. From genes to protein mechanics on a chip. Nat. Methods 11, 1127–1130. http://dx.doi.org/10.1038/nmeth.3099.

Pfreundschuh, M., Alsteens, D., Hilbert, M., Steinmetz, M.O., Müller, D.J., 2014. Localizing chemical groups while imaging single native proteins by high-resolution atomic force microscopy. Nano Lett. 14, 2957–2964. http://dx.doi.org/10.1021/nl5012905.

Pfreundschuh, M., Alsteens, D., Wieneke, R., Zhang, C., Coughlin, S.R., Tampé, R., Kobilka, B.K., Müller, D.J., 2015. Identifying and quantifying two ligand-binding sites while imaging native human membrane receptors by AFM. Nat. Commun. 6, 1–7. http://dx.doi.org/10.1038/ncomms9857.

Pippig, D.A., Baumann, F., Strackharn, M., Aschenbrenner, D., Gaub, H.E., 2014. Protein–DNA chimeras for nano assembly. ACS Nano 8, 6551–6555. http://dx.doi.org/10.1021/nn501644w.

Popa, I., Berkovich, R., Alegre-Cebollada, J., Badilla, C.L., Rivas-Pardo, J.A., Taniguchi, Y., Kawakami, M., Fernandez, J.M., 2013. Nanomechanics of HaloTag tethers. J. Am. Chem. Soc. 135, 12762–12771. http://dx.doi.org/10.1021/ja4056382.

Preiner, J., Kodera, N., Tang, J., Ebner, A., Brameshuber, M., Blaas, D., Gelbmann, N., Gruber, H.J., Ando, T., Hinterdorfer, P., 2014. IgGs are made for walking on bacterial and viral surfaces. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms5394.

Puchner, E.M., Franzen, G., Gautel, M., Gaub, H.E., 2008. Comparing proteins by their unfolding pattern. Biophys. J. 95, 426–434. http://dx.doi.org/10.1529/biophysj.108.129999.

Raab, A., Hansma, H.G., Badt, D., Smith-Gill, S.J., 1999. Antibody recognition imaging by force microscopy. Nature. http://dx.doi.org/10.1038/12898.

Radmacher, M., Tillmann, R.W., Fritz, M., Gaub, H.E., 1992. From molecules to cells – imaging soft samples with the atomic force microscope. Science 257, 1900–1905. http://dx.doi.org/10.1126/science.1411505.

Radmacher, M., Fritz, M., Kacher, C.M., Cleveland, J.P., Hansma, P.K., 1996. Measuring the viscoelastic properties of human platelets with the atomic force microscope. Biophys. J. 70, 556–567. http://dx.doi.org/10.1016/S0006-3495(96)79602-9.

Rico, F., Moy, V.T., 2007. Energy landscape roughness of the streptavidin–biotin interaction. J. Mol. Recogn. 20, 495–501. http://dx.doi.org/10.1002/jmr.841.

Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M., Scheuring, S., 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. Science 342, 741–743. http://dx.doi.org/10.1126/science.1239764.

Rief, M., Grubmüller, H., 2002. Force spectroscopy of single biomolecules. Chemphyschem 3, 255–261. http://dx.doi.org/10.1002/1439-7641(20020315)3:3<255::AID-CPHC255>3.0.CO;2-M.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., Gaub, H.E., 1997a. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276, 1109–1112. http://dx.doi.org/10.1126/science.276.5315.1109.

Rief, M., Oesterhelt, F., Heymann, B., Gaub, H.E., 1997b. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. Science 275, 1295–1297. http://dx.doi.org/10.1126/science.275.5304.1295.

Schmidt, S.W., Filippov, P., Kersch, A., Beyer, M.K., Clausen-Schaumann, H., 2012. Single-molecule force-clamp experiments reveal kinetics of mechanically activated silyl ester hydrolysis. ACS Nano 6, 1314–1321. http://dx.doi.org/10.1021/nn204111w.

Schoeler, C., Malinowska, K.H., Bernardi, R.C., Milles, L.F., Jobst, M.A., Durner, E., Ott, W., Fried, D.B., Bayer, E.A., Schulten, K., Gaub, H.E., Nash, M.A., 2014. Ultrastable cellulosome–adhesion complex tightens under load. Nat. Commun. 5, 1–8. http://dx.doi.org/10.1038/ncomms6635.

Schoeler, C., Bernardi, R.C., Malinowska, K.H., Durner, E., Ott, W., Bayer, E.A., Schulten, K., Nash, M.A., Gaub, H.E., 2015. Mapping mechanical force propagation through biomolecular complexes. Nano Lett. 15, 7370–7376. http://dx.doi.org/10.1021/acs.nanolett.5b02727.

Schuler, B., Hofmann, H., 2013. Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. Curr. Opin. Struct. Biol. 23, 36–47. http://dx.doi.org/10.1016/j.sbi.2012.10.008.

Schwesinger, F., Ros, R., Strunz, T., Anselmetti, D., Guntherodt, H.-J., Honegger, A., Jermutus, L., Tiefenauer, L., Plückthun, A., 2000. Unbinding forces of single antibody–antigen complexes correlate with their thermal dissociation rates. Proc. Natl. Acad. Sci. USA 97, 9972–9977. http://dx.doi.org/10.1073/pnas.97.18.9972.

Sirbuly, D.J., Friddle, R.W., Villanueva, J., Huang, Q., 2015. Nanomechanical force transducers for biomolecular and intracellular measurements: is there room to shrink and why do it? Rep. Prog. Phys. 1–22. http://dx.doi.org/10.1088/0034-4885/78/2/024101.

Smith, S.B., Finzi, L., Bustamante, C., 1992. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. Science 258, 1122–1126. http://dx.doi.org/10.1126/science.1439819.

Stahl, S.W., Nash, M.A., Fried, D.B., Slutzki, M., Barak, Y., Bayer, E.A., Gaub, H.E., 2012. Single-molecule dissection of the high-affinity cohesin–dockerin complex. Proc. Natl. Acad. Sci. USA 109, 20431–20436. http://dx.doi.org/10.1073/pnas.1211929109.

Sułkowska, J.I., Cieplak, M., 2007. Mechanical stretching of proteins—a theoretical survey of the protein data bank. J. Phys. Condens. Matter 19, 283201. http://dx.doi.org/10.1088/0953-8984/19/28/283201.

Svoboda, K., Schmidt, C.F., Schnapp, B.J., Block, S.M., 1993. Direct observation of kinesin stepping by optical trapping interferometry. Nature 365, 721–727. http://dx.doi.org/10.1038/365721a0.

Thoma, J., Burmann, B.M., Hiller, S., Müller, D.J., 2015. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. Nat. Struct. Mol. Biol. 22, 795–802. http://dx.doi.org/10.1038/nsmb.3087.

Tsukasaki, Y., Kitamura, K., Shimizu, K., Iwane, A.H., Takai, Y., Yanagida, T., 2007. Role of multiple bonds between the single cell adhesion molecules, nectin and cadherin, revealed by high sensitive force measurements. J. Mol. Biol. 367, 996–1006. http://dx.doi.org/10.1016/j.jmb.2006.12.022.

Valbuena, A., Oroz, J., Hervas, R., Manuel Vera, A., Rodriguez, D., Menendez, M., Sulkowska, J.I., Cieplak, M., Carrion-Vazquez, M., 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc. Natl. Acad. Sci. USA 106, 13791–13796. http://dx.doi.org/10.1073/pnas.0813093106.

Verbelen, C., Gruber, H.J., Dufrêne, Y.F., 2007. The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. J. Mol. Recogn. 20, 490–494. http://dx.doi.org/10.1002/jmr.833.

Viani, M.B., Schaffer, T.E., Chand, A., Rief, M., Gaub, H.E., Hansma, P.K., 1999a. Small cantilevers for force spectroscopy of single molecules. J. Appl. Phys. 86, 2258–2262. http://dx.doi.org/10.1063/1.371039.

Viani, M.B., Schaffer, T.E., Paloczi, G.T., Pietrasanta, L.I., Smith, B.L., Thompson, J.B., Richter, M., Rief, M., Gaub, H.E., Plaxco, K.W., Cleland, A.N., Hansma, H.G., Hansma, P.K., 1999b. Fast imaging and fast force spectroscopy of single biopolymers with a new atomic force microscope designed for small cantilevers. Rev. Sci. Instrum. 70, 4300–4303. http://dx.doi.org/10.1063/1.1150069.

Walder, R., Paik, D.H., Bull, M.S., Sauer, C., Perkins, T.T., 2015. Ultrastable measurement platform: sub-nm drift over hours in 3D at room temperature. Opt. Express 23, 16554–16564. http://dx.doi.org/10.1364/OE.23.016554.

Wildling, L., Rankl, C., Haselgrübler, T., Gruber, H.J., Holy, M., Newman, A.H., Zou, M.-F., Zhu, R., Freissmuth, M., Sitte, H.H., Hinterdorfer, P., 2012. Probing binding pocket of serotonin transporter by single molecular force spectroscopy on living cells. J. Biol. Chem. 287, 105–113. http://dx.doi.org/10.1074/jbc.M111.304873.

Wong, J., Chilkoti, A., Moy, V.T., 1999. Direct force measurements of the streptavidin–biotin interaction. Biomol. Eng. 16, 45–55. http://dx.doi.org/10.1016/S1050-3862(99)00035-2.

Woodside, M.T., Block, S.M., 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu. Rev. Biophys. 43, 19–39. http://dx.doi.org/10.1146/annurev-biophys-051013-022754.

Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R., Walsh, C.T., 2005. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc. Natl. Acad. Sci. USA 102, 15815–15820. http://dx.doi.org/10.1073/pnas.0507705102.

Yin, J., Lin, A.J., Golan, D.E., Walsh, C.T., 2006. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. Nat. Protocols 1, 280–285. http://dx.doi.org/10.1038/nprot.2006.43.

Yu, H., Liu, X., Neupane, K., Gupta, A.N., Brigley, A.M., Solanki, A., Sosova, I., Woodside, M.T., 2012. Direct observation of multiple misfolding pathways in a single prion protein molecule. Proc. Natl. Acad. Sci. USA 109, 5283–5288. http://dx.doi.org/10.1073/pnas.1107736109.

Yu, H., Dee, D.R., Liu, X., Brigley, A.M., Sosova, I., Woodside, M.T., 2015. Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. Proc. Natl. Acad. Sci. USA 112, 8308–8313. http://dx.doi.org/10.1073/pnas.1419197112.

Yuan, C., Chen, A., Kolb, P., Moy, V.T., 2000. Energy landscape of streptavidin–biotin complexes measured by atomic force microscopy. Biochemistry 39, 10219–10223. http://dx.doi.org/10.1021/bi992715o.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., Howarth, M., 2012. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc. Natl. Acad. Sci. USA 109, 4347–4348. http://dx.doi.org/10.1073/pnas.1115485109.

Zhang, S., Aslan, H., Besenbacher, F., Dong, M., 2014. Quantitative biomolecular imaging by dynamic nanomechanical mapping. Chem. Soc. Rev. 43, 7412–7429. http://dx.doi.org/10.1039/C4CS00176A.

Zhou, J., Schweikhard, V., Block, S.M., 2013. Single-molecule studies of RNAPII elongation. Biochem. Biophys. Acta 1829, 29–38. http://dx.doi.org/10.1016/j.bbagrm.2012.08.006.

Zimmermann, J.L., Nicolaus, T., Neuert, G., Blank, K., 2010. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. Nat. Protocols 5, 975–985. http://dx.doi.org/10.1038/nprot.2010.49.

Žoldák, G., Stigler, J., Pelz, B., Li, H., 2013. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. Proc. Natl. Acad. Sci. USA 110, 18156–18161. http://dx.doi.org/10.1073/pnas.1311495110.

Chapter 4

Biomechanics of the Cellulosome

4.1 Summary

The cellulosome is an excellent model system to study the mechanobiology of single molecule interactions. It is an extracellular multiprotein complex, relying on multidomain framework proteins, called scaffoldins. Scaffoldins contain several cohesin domains and often a carbohydrate binding module (CBM). Cohesins recognize dockerin domains, which are fused to the catalytic subunits of enzymes. Thus different enzymes are arranged in close proximity to each other on the scaffoldin through cohesin:dockerin interactions. The CBM ensures the localization of the enzyme ensemble to their substrate. The scaffoldin is connected to its secreting host cell with an orthogonal cohesin:dockerin interaction (type II or III) than the one recruiting enzymes (type I), enabling a hierarchical assembly of host cell, scaffoldins and enzymes [15]. The multitude of different non cross-reacting, high affinity interactions make it an interesting object to study with single molecule force spectroscopy (SMFS).
Moreover, cellulolytic organisms live in turbulent environments (*i.e.* hot springs, rumen of cows, or digestive systems of humans) [126], rendering cellulosomes especially attractive for biomechanical studies.
In **associated publication P2** a unique type III interaction, between the cell and scaffoldin from *Ruminococcus flavefaciens*, was probed. It was hypothesized that an interaction between the cell and scaffoldin might resist high forces because of its anchoring function. Indeed, with complex rupture forces of 650 pN one of the highest single complex rupture forces could be observed. All-atom simulations revealed that the high forces are caused by an increased binding interface area upon force loading.
Associated publication P3 continued working with the type III cohesin:dockerin interaction and compared native with non-native pulling geometries. The cohesin domain is natively located at the N-terminus, when the domain is moved to the C-terminus of a pulling experiment unfolding forces drop. Again molecular dynamic simulations explained the molecular mechanism. In the non-native configuration, force propagates almost on a straight line through the protein, dissipating less mechanical load than in the native geometry.
SMFS studies can also resolve different binding modes of the same receptor:ligand interaction, as shown in **associated publication P4:** A highly symmetric type I dockerin from *Clostridium thermocellum* can bind the corresponding cohesin in two modes, that are rotated by 180°. So far evidence for the dual binding mode only came from crystal structures of mutants, that showed one or the other binding mode, but never both [65], [127], [128]. The atomic force microscopy (AFM) **study P4** showed that the wildtype receptor:ligand interaction exerts two binding modes with two different unbinding forces.
SMFS is also able to probe the mechanical stability of protein domains. In **associated publication P5** all cohesin domains of one scaffoldin from *Acetivibrio cellulolyticus*

were probed. The scaffoldin connects the host cell with cellulose fibers *via* its CBM. The CBM is located internally in the scaffoldin, so that some cohesins are located between cell and cellulose (bridging) and some are free (hanging). Hence, mechanical stability of bridging cohesins should be higher than the hanging ones. AFM measurements confirmed this hypothesis. Furthermore, crucial amino acids, of the more stable cohesins could be identified with all-atom simulations and transferred to the weakest fold, increasing its mechanostabilty.
AFM SMFS is a valuable tool to probe biomechanics of a system. Insights can be transfered directly to new applications, *e.g.* using the type III cohesin:dockerin as a new binding handle in SMFS experiments.

4.2 Associated Publication P2

Ultrastable cellulosome-adhesion complex tightens under load

by

Constantin Schoeler*, Klara H. Malinowska*, Rafael C. Bernardi,

Lukas F. Milles, Markus A. Jobst, Ellis Durner, Wolfgang Ott,

Daniel B. Fried, Edward A. Bayer, Klaus Schulten,

Hermann E. Gaub and Michael A. Nash

*(*contributed equally)*

published in

Nature Communications, 5, 1-8, (2014)

nature COMMUNICATIONS

ARTICLE

Received 25 Jun 2014 | Accepted 22 Oct 2014 | Published 8 Dec 2014

DOI: 10.1038/ncomms6635 OPEN

Ultrastable cellulosome-adhesion complex tightens under load

Constantin Schoeler[1,*], Klara H. Malinowska[1,*], Rafael C. Bernardi[2], Lukas F. Milles[1], Markus A. Jobst[1], Ellis Durner[1], Wolfgang Ott[1], Daniel B. Fried[3], Edward A. Bayer[3], Klaus Schulten[2,4], Hermann E. Gaub[1] & Michael A. Nash[1]

Challenging environments have guided nature in the development of ultrastable protein complexes. Specialized bacteria produce discrete multi-component protein networks called cellulosomes to effectively digest lignocellulosic biomass. While network assembly is enabled by protein interactions with commonplace affinities, we show that certain cellulosomal ligand–receptor interactions exhibit extreme resistance to applied force. Here, we characterize the ligand–receptor complex responsible for substrate anchoring in the *Ruminococcus flavefaciens* cellulosome using single-molecule force spectroscopy and steered molecular dynamics simulations. The complex withstands forces of 600–750 pN, making it one of the strongest bimolecular interactions reported, equivalent to half the mechanical strength of a covalent bond. Our findings demonstrate force activation and inter-domain stabilization of the complex, and suggest that certain network components serve as mechanical effectors for maintaining network integrity. This detailed understanding of cellulosomal network components may help in the development of biocatalysts for production of fuels and chemicals from renewable plant-derived biomass.

[1] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany. [2] Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3] Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel. [4] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.A.N. (email: michael.nash@lmu.de).

Cellulosomes are protein networks designed by nature to degrade lignocellulosic biomass[1]. These networks comprise intricate assemblies of conserved subunits including catalytic domains, scaffold proteins, carbohydrate binding modules (CBMs), cohesins (Cohs), dockerins (Docs) and X-modules (XMods) of unknown function. Coh:Doc pairs form complexes with high affinity and specificity[2], and provide connectivity to a myriad of cellulosomal networks with varying Coh:Doc network topology[3–5]. The most intricate cellulosome known to date is produced by *Ruminococcus flavefaciens* (*R.f.*)[6,7] and contains several primary and secondary scaffolds along with over 220 Doc-bearing protein subunits[8].

The importance of cellulolytic enzymes for the production of renewable fuels and chemicals from biomass has highlighted an urgent need for improved fundamental understanding of how cellulosomal networks achieve their impressive catalytic activity[9]. Two of the mechanisms known to increase the catalytic activity of cellulosomes are proximity and targeting effects[10]. Proximity refers to the high local concentration of enzymes afforded by incorporation into nanoscale networks, while targeting refers to specific binding of cellulosomes to substrates. Protein scaffolds and CBM domains are both critical in this context as they mediate interactions between comparatively large bacterial cells and cellulose particles. As many cellulosomal habitats (for example, cow rumen) exhibit strong flow gradients, shear forces will accordingly stress bridging scaffold components mechanically *in vivo*. Protein modules located at stressed positions within these networks should therefore be preselected for high mechanostability. However, thus far very few studies on the mechanics of carbohydrate-active proteins or cellulosomal network components have been reported[11].

In the present study we sought to identify cellulosomal network junctions with maximal mechanical stability. We chose an XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose in the rumen. The complex links the *R. flavefaciens* cell wall to the cellulose substrate via two CBM domains located at the N-terminus of the CttA scaffold, as shown in Fig. 1a. The crystal structure of the complex solved by X-ray crystallography[12] is shown in Fig. 1b. XMod-Doc tandem dyads such as this one are a common feature in cellulosomal networks. Bulk biochemical assays on XMod-Docs have demonstrated that XMods improve Doc solubility and increase biochemical affinity of Doc:Coh complex formation[13]. Crystallographic studies conducted on XMod-Doc:Coh complexes have revealed direct contacts between XMods and their adjacent Docs[12,14]. In addition, many XMods (for example, PDB 2B59, 1EHX, 3PDD) have high β-strand content and fold with N- and C-termini at opposite ends of the molecule, suggestive of robust mechanical clamp motifs at work[15,16]. These observations all suggest a mechanical role for XMods. Here we perform AFM single-molecule force spectroscopy experiments and steered molecular dynamics simulations to understand the mechanostability of the XMod-Doc:Coh cellulosomal ligand–receptor complex. We conclude that the high mechanostability we observe originates from molecular mechanisms, including stabilization of Doc by the adjacent XMod domain and catch bond behaviour that causes the complex to increase in contact area on application of force.

Results and Discussion

Single-molecule experiments. We performed single-molecule force spectroscopy (SMFS) experiments with an atomic force miscroscope (AFM) to probe the mechanical dissociation of XMod-Doc:Coh. Xylanase (Xyn) and CBM fusion domains on the XMod-Doc and Coh modules, respectively, provided identifiable unfolding patterns permitting screening of large data sets of force-distance curves[17–19]. Engineered cysteines and/or peptide tags on the CBM and Xyn marker domains were used to covalently immobilize the binding partners in a site-specific manner to an AFM cantilever or cover glass via poly(ethylene glycol) (PEG) linkers. The pulling configuration with Coh-CBM immobilized on the cantilever is referred to as configuration I, as shown in Fig. 1c. The reverse configuration with Coh-CBM on the cover glass is referred to as configuration II. In a typical

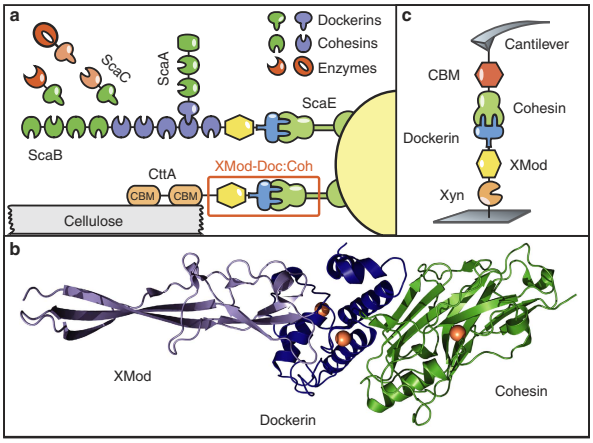


Figure 1 | System overview. (**a**) Schematic of selected components of the *R. flavefaciens* cellulosome. The investigated XMod-Doc:Coh complex responsible for maintaining bacterial adhesion to cellulose is highlighted in orange. (**b**) Crystal structure of the XMod-Doc:Coh complex. Ca^{2+} ions are shown as orange spheres. (**c**) Depiction of experimental pulling configuration I, with Coh-CBM attached to the cantilever tip and Xyn-XMod-Doc attached to the glass surface.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms6635 ARTICLE

experimental run we collected about 50,000 force extension traces from a single cantilever. We note that the molecules immobilized on the cantilever and glass surfaces were stable over thousands of pulling cycles.

We sorted the data by first searching for contour length increments that matched our specific xylanase and CBM fingerprint domains. After identifying these specific traces (Fig. 2a), we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history. To assign protein subdomains to the observed unfolding patterns, we transformed the data into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching (QM-FRC, Supplementary Note 1, Supplementary Fig. 1)[20,21]. The fit parameter-free QM-FRC model describes protein stretching at forces >200 pN more accurately than the commonly used worm-like chain (WLC) model[20,22]. The resulting contour length histogram is shown in Fig. 2b. Peak-to-peak distances in the histogram represent contour length increments of unfolded protein domains. Assuming a length per stretched amino acid of 0.365 nm and accounting for the folded length of each subdomain, we compared the observed increments to the polypeptide lengths of individual subdomains of the Xyn-XMod-Doc and Coh-CBM fusion proteins. Details on contour length estimates and domain assignments are shown in Supplementary Table 1.

Unfolding patterns in configuration I showed PEG stretching followed by a three-peaked Xyn fingerprint (Fig. 1a, top trace, green), which added 90 nm of contour length to the system. Xyn unfolding was followed by CBM unfolding at ~150 pN with 55 nm of contour length added. Finally, the XMod-Doc:Coh complex dissociated at an ultra-high rupture force of ~600 pN. The loading rate dependence of the final rupture event for curves of subtype 1 is plotted in Fig. 2c (blue). The measured complex rupture force distributions are shown in Supplementary Fig. 2.

Less frequently (35–40% of traces) we observed a two-step dissociation process wherein the XMod unfolded before Doc:Coh rupture as shown in Fig. 2a (middle trace, orange). In these cases, the final dissociation exhibited a much lower rupture force (~300 pN) than the preceding XMod unfolding peak, indicating the strengthening effect of XMod was lost, and XMod was no longer able to protect the complex from dissociation at high force. The loading rate dependency of Doc:Coh rupture occurring immediately following XMod unfolding is shown in Fig. 2c (grey).

In configuration II (Fig. 2a, bottom trace), with the Xyn-XMod-Doc attached to the cantilever, the xylanase fingerprint was lost after the first few force extension traces acquired in the data set. This indicated the Xyn domain did not refold within the timescale of the experiment once unfolded, consistent with prior work[17,18]. CBM and XMod unfolding events were observed repeatedly throughout the series of acquired force traces in both configurations I and II, indicating these domains were able to refold while attached to the cantilever over the course of the experiment.

We employed the Bell-Evans model[23] (Supplementary Note 2) to analyse the final rupture of the complex through the effective distance to the transition state (Δx) and the natural off-rate (k_{off}). The fits to the model yielded values of $\Delta x = 0.13$ nm and $k_{off} = 7.3 \times 10^{-7}\,s^{-1}$ for an intact XMod, and $\Delta x = 0.19$ nm and $k_{off} = 4.7 \times 10^{-4}\,s^{-1}$ for the 'shielded' rupture following XMod unfolding (Fig. 2c). These values indicate that the distance to the transition state is increased following XMod unfolding, reflecting an overall softening of the binding interface. Distances to the transition state observed for other ligand–receptor pairs are typically on the order of ~0.7 nm (ref. 17). The extremely short Δx of 0.13 nm observed here suggests that mechanical unbinding for this complex is highly coordinated. We further analysed the unfolding of XMod in the Bell-Evans picture and found values of $\Delta x = 0.15$ and $k_{off} = 2.6 \times 10^{-6}\,s^{-1}$. The loading

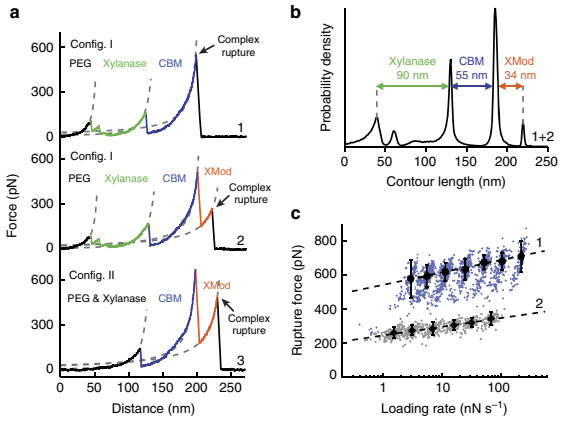


Figure 2 | Experimental SMFS unfolding traces. (**a**) Unfolding fingerprints from pulling configuration I (curves 1 & 2) and configuration II (curve 3). The QM-FRC model (dashed lines) was used to estimate the contour lengths of the unfolded modules. (**b**) Contour length histogram obtained from 127 force extension traces (Config. I). The peak-to-peak increments correspond to Xyn, CBM and XMod amino-acid sequence lengths. (**c**) Dynamic force spectra for the final Doc:Coh complex rupture peaks obtained from 2,122 force-extension traces. The blue points show Doc:Coh ruptures that occurred with an intact XMod, while grey points show ruptures immediately following XMod unfolding. Black circles and diamonds represent the most probable rupture force/loading rate obtained by Gaussian fitting at each pulling speed. Error bars are ±1 s.d. Dashed lines are least square fits to the Bell-Evans model.

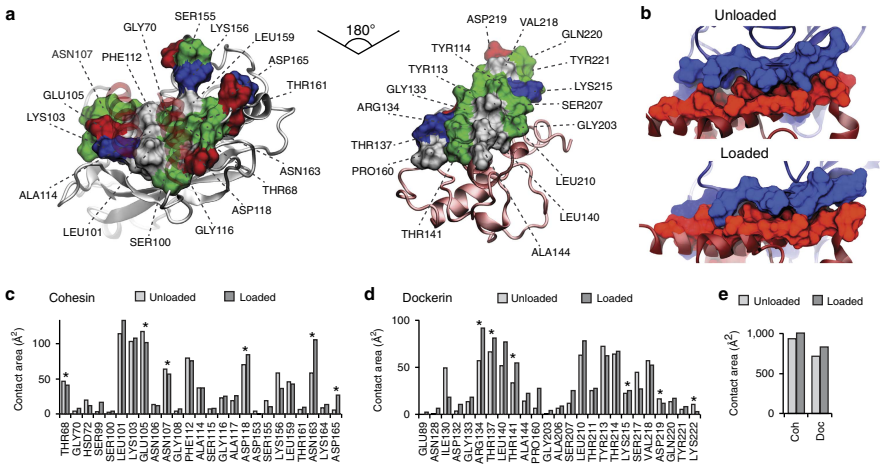


Figure 3 | Analysis of binding interface and catch bond mechanism from SMD. (**a**) Surface plots for the main interacting residues of Coh (left) and Doc (right). Hydrophobic residues are shown in grey, polar residues in green, and negative and positive residues in red and blue, respectively. Both Coh and Doc exhibit a hydrophobic patch in the centre of the binding surface that is surrounded by polar and charged residues. (**b**) Rearrangement of binding residues of Coh (blue) and Doc (red) under force. Following mechanical loading, an interdigitated complex is formed that resembles teeth of a zipper. (**c**,**d**) Surface contact area of interacting residues of Coh (c) and Doc (d) in the absence and presence of force. Residues forming prevalent hydrogen bonds are indicated with stars. (**e**) Total contact surface area of Coh and Doc in unloaded and loaded conformations.

rate dependence for this unfolding event is shown in Supplementary Fig. 3.

The exceptionally high rupture forces measured experimentally (Fig. 2) are hugely disproportionate to the XMod-Doc:Coh biochemical affinity, which at $K_D \sim 20$ nM (ref. 12) is comparable to typical antibody–antigen interactions. Antibody–antigen interactions, however, will rupture at only ~ 60 pN at similar loading rates[24], while bimolecular complexes found in muscle exposed to mechanical loading *in vivo* will rupture at ~ 140 pN (ref. 25). Trimeric titin–telethonin complexes also found in muscle exhibit unfolding forces around 700 pN (ref. 26), while Ig domains from cardiac titin will unfold at ~ 200 pN (ref. 27). The XMod-Doc:Coh ruptures reported here fell in a range from 600 to 750 pN at loading rates ranging from 10 to 100 nN s^{-1}. At around half the rupture force of a covalent gold-thiol bond[28], these bimolecular protein rupture forces are, to the best of our knowledge, among the highest of their kind ever reported. The covalent bonds in this system are primarily peptide bonds in the proteins and C-C and C-O bonds in the PEG linkers. These are significantly more mechanically stable than the quoted gold-thiol bond rupture force (~ 1.2 nN) (ref. 29) and fall in a rupture force range > 2.5 nN at similar loading rates. Therefore, breakage of covalent linkages under our experimental conditions is highly unlikely. We note that the high mechanostability observed here is not the result of fusing the proteins to the CBM or Xyn domains. The covalent linkages and pulling geometry are consistent with the wild-type complex and its dissociation pathway. *In vivo*, the Coh is anchored to the peptidoglycan cell wall through its C-terminal sortase motif. The XMod–Doc is attached to the cellulose substrate through two N-terminal CBM domains. By pulling the XMod–Doc through an N-terminal Xyn fusion domain, and the Coh through a C-terminal CBM, we established an experimental pulling geometry that matches loading of the complex *in vivo*. This pulling geometry was also used in all simulations. The discontinuity between its commonplace biochemical affinity and remarkable resistance to applied force illustrates how this complex is primed for mechanical stability and highlights differences in the unbinding pathway between dissociation at equilibrium and dissociation induced mechanically along a defined pulling coordinate.

Steered molecular dynamics. To elucidate the molecular mechanisms at play that enable this extreme mechanostability, we carried out all-atom steered molecular dynamics (SMD) simulations. The Xyn and CBM domains were not modelled to keep the simulated system small and reduce the usage of computational resources. This approximation was reasonable as we have no indication that these domains significantly affect the XMod–Doc:Coh binding strength[30]. After equilibrating the crystal structure[12], the N-terminus of XMod–Doc was harmonically restrained while the C-terminus of Coh was pulled away at constant speed. The force applied to the harmonic pulling spring was stored at each time step. We tested pulling speeds of 0.25, 0.625 and 1.25 Å ns^{-1}, and note that the slowest simulated pulling speed was $\sim 4{,}000$ times faster than our fastest experimental pulling speed of 6.4 μm s^{-1}. This difference is considered not to affect the force profile, but it is known to account for the scale difference in force measured by SMD and AFM[31,32].

SMD results showed the force increased with distance until the complex ruptured for all simulations. At the slowest pulling speed of 0.25 Å ns^{-1} the rupture occurred at a peak force of ~ 900 pN, as shown in Supplementary Fig. 4 and Supplementary Movie 1. We analysed the progression and prevalence of hydrogen bonded contacts between the XMod–Doc and Coh domains to identify

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms6635 ARTICLE

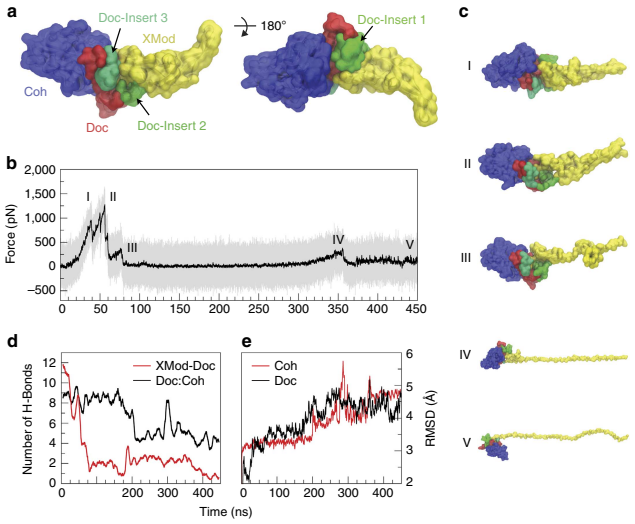


Figure 4 | SMD shows unfolding of XMod destabilizes Doc:Coh binding interface. XMod was unfolded by moving the harmonic restraint to the C terminus of XMod while the N terminus was moved at 0.625 Å ns^{-1}. (**a**) Surface representation of XMod-Doc:Coh complex with Doc insert sequences. Coh is shown in blue, Doc in red and green (inserts), and XMod in yellow. (**b**) Force time trace of XMod unfolding. The domain starts to unfold in several substeps starting at ~400 pN. Snapshots at different time steps are labelled I-V and are shown in (**c**). Steps IV and V are shown at smaller scale. (**d**) Average number of hydrogen bonds between Doc:Coh (black) and XMod-Doc (red). XMod-Doc contact is dominated by the insert sequences 1–3. (**e**) Root mean squared deviation (RMSD) of Doc (black) and Coh (red).

key residues in contact throughout the entire rupture process and particularly immediately before rupture. These residues are presented in Fig. 3a,c,d and Supplementary Figs 5,6. The simulation results clearly reproduced key hydrogen bonding contacts previously identified[12] as important for Doc:Coh recognition (Supplementary Fig. 5).

The main interacting residues are shown in Fig. 3a,b. Both Coh and Doc exhibit a binding interface consisting of a hydrophobic centre (grey) surrounded by a ring of polar (green) and charged residues (blue, positive; red, negative). This residue pattern suggests the hydrophilic side chains protect the interior hydrophobic core from attack by water molecules, compensating for the flat binding interface that lacks a deep pocket. The geometry suggests a penalty to unbinding that stabilizes the bound state. Further, we analysed the contact surface areas of interacting residues (Fig. 3b–e). The total contact area was found to increase due to rearrangement of the interacting residues when the complex is mechanically stressed, as shown in Fig. 3e and Supplementary Movie 2. Doc residues in the simulated binding interface clamped down on Coh residues upon mechanical loading, resulting in increased stability and decreased accessibility of water into the hydrophobic core of the bound complex (Fig. 3b). These results suggest that a catch bond mechanism is responsible for the remarkable stability[33] under force and provide a molecular mechanism which the XMod–Doc:Coh complex uses to summon mechanical strength when needed, while still allowing relatively fast assembly and disassembly of the complex at equilibrium. The residues that increase most in contact area (Fig. 3c,d) present promising candidates for future mutagenesis studies.

Among the 223 Doc sequences from *R. flavefaciens*, six subfamilies have been explicitly identified using bioinformatics approaches[8]. The XMod–Doc investigated here belongs to the 40-member Doc family 4a. A conserved feature of these Doc modules is the presence of three sequence inserts that interrupt the conserved duplicated F-hand motif Doc structure. In our system, these Doc sequence inserts make direct contacts with XMod in the crystallized complex (Fig. 1) and suggest an interaction between XMod and Doc that could potentially propagate to the Doc:Coh binding interface. To test this, an independent simulation was performed to unfold XMod (Fig. 4). The harmonic restraint was moved to the C-terminus of XMod so that force was applied from the N- to C-terminus of XMod only, while leaving Doc and Coh unrestrained. The results (Fig. 4b) showed XMod unfolded at forces slightly higher than but similar to the XMod–Doc:Coh complex rupture force determined from the standard simulation at the same pulling speed. This suggested XMod unfolding before Doc:Coh rupture was not probable, but could be observed on occasion due to the stochastic nature of domain unfolding. This was consistent with experiments where XMod unfolding was observed in ~35–40% of traces. Furthermore, analysis of the H-bonding between Doc and XMod (Fig. 4d, red) indicated loss of contact as XMod unfolded, dominated by contact loss between the three Doc insert sequences and XMod. Interestingly, XMod unfolding clearly led to a decrease in H-bonding between Doc and Coh at a later stage (~200 ns) well after XMod had lost most of its contact with Doc, even though no force was being applied across the Doc:Coh binding interface. This provided evidence for direct stabilization of the Doc:Coh binding interface by XMod.

As shown in Fig. 4e, the root mean squared deviation (RMSD) of Doc increased throughout the simulation as XMod unfolded. Coh RMSD remained stable until it started to lose H-bonds with Doc. Taken together this suggests that, as XMod unfolded, Coh and Doc became more mobile and lost interaction strength, potentially explaining the increase in Δx from 0.13 to 0.19 nm on unfolding of XMod in the experimental data sets. Apparently the XMod is able to directly stabilize the Doc:Coh interface, presumably through contact with Doc insert sequences that then propagate this stabilizing effect to the Doc:Coh binding interface.

In summary, we investigated an ultrastable XMod-Doc:Coh complex involved in bacterial adhesion to cellulose. While previously the role of XMod functioning in tandem XMod-Doc dyads was unclear[12,14], we show that XMod serves as a mechanical stabilizer and force-shielding effector subdomain in the ultrastable ligand–receptor complex. The Doc:Coh complex presented here exhibits one of the most mechanically robust protein–protein interactions reported thus far, and points towards new mechanically stable artificial multi-component biocatalysts for industrial applications, including production of second-generation biofuels.

Methods

Site-directed mutagenesis. Site-directed mutagenesis of *R. flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus*[12] were subjected to QuikChange mutagenesis[34] to install the following mutations: A2C in the CBM and T129C in the xylanase, respectively.

For the construction of the native configuration of the CohE-CBM A2C fusion protein Gibson assembly[35] was used. For further analysis CohE-CBM A2C was modified with a QuikChange PCR[36] to replace the two cysteins (C2 and C63) in the protein with alanine and serine (C2A and C63S). All mutagenesis products were confirmed by DNA sequencing analysis.

The XynT6-XDoc T129C was constructed using the following primers:
5′-acaaggaaggtaagccaatggttaatgaatgcgatccagtgaaacgtgaac-3′
5′-gttcacgtttcactggatcgcattcattaaccattggcttaccttcctttgt-3′

The CBM-CohE A2C was constructed using the following primers:
5′-ttaactttaagaaggagatataccatgtgcaatacaccggtatcaggcaatttgaag-3′
5′-cttcaaattgcctgataccggtgtattgcacatggtatatctccttcttaaagttaa-3′

The CohE-CBM C2A C63S was constructed using the following phosphorylated primers:
5′-ccgaatgccatggccaatacaccgg-3′
5′-cagaccttctggagtgaccatgctgc-3′

Expression and purification of Xyn-XMod-Doc. The T129C Xyn-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with tris-buffered saline (TBS) buffer containing 20 mM imidazole and 2 mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialysed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at − 20 °C. The concentrations of the protein stock solutions were determined to be ∼5 mg ml^{-1} by absorption spectrophotometry.

Expression and purification of Coh-CBM. The Coh-CBM C2A, C63S fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicol containing ZYM-5052 media[37] overnight at 22 °C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at − 80 °C. The concentrations of the protein stock solutions were determined to be ∼5 mg ml^{-1} by absorption spectrophotometry.

Sample preparation. In sample preparation and single-molecule measurements calcium supplemented TBS buffer (Ca-TBS) was used (25 mM TRIS, 72 mM NaCl, 1 mM $CaCl_2$, pH 7.2). Cantilevers and cover glasses were functionalized according to previously published protocols[18,38]. In brief, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyn-XMod-Doc proteins were reduced for 2 h at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with Ca-TBS measurement buffer, centrifuged at 850 r.c.f. for 3 min, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with Ca-TBS to remove unbound proteins and stored under Ca-TBS before force spectroscopy measurements. Site-specific immobilization of the Coh-CBM-ybbR fusion proteins to previously PEGylated cantilevers or coverglasses was carried out according to previously published protocols[39]. In brief, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) for 1 h at room temperature. Levers or surfaces were then rinsed with Ca-TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 h at room 37°. Finally, surfaces or levers were subjected to a final rinse with Ca-TBS and stored under Ca-TBS before measurement.

Single-molecule force spectroscopy measurements. SMFS measurements were performed on a custom built AFM[40] controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise/equipartition method[41]. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2 to 6.4 μm s^{-1}. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18–24 h. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw data sets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single-molecule interactions. We filtered the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: Xyn (∼89 nm) and CBM (∼56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

Data analysis. Data were analysed using previously published protocols[17,18,22]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11$ nm connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. Detailed description of the contour length transformation can be found in Supplementary Note 1 and Supplementary Fig. 1.

For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs time trace in the immediate vicinity before the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying Gaussian fits to histograms of rupture forces and loading rates at each pulling speed.

Molecular dynamics simulations. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97 Å resolution and is available at the protein data bank (PDB:4IU3). A protonation analysis performed in VMD[42] did not suggest any extra protonation and all the amino-acid residues were simulated with standard protonation states. The system was then solvated, keeping also the water molecules present in the crystal structure, and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter ions, which were randomly arranged in the solvent. Two other systems, based on the aforementioned one, were created using a similar salt concentration to the one used in the experiments (75 mM of NaCl). This additional salt caused little or no change in SMD results. The overall number of atoms included in MD simulations varied from 300,000 in the majority of the simulations to 580,000 for the unfolding of the X-Mod.

The MD simulations in the present study were performed employing the NAMD molecular dynamics package[43,44]. The CHARMM36 force field[45,46] along with the TIP3 water model[47] was used to describe all systems. The simulations were done assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were

treated using the particle-mesh Ewald (PME)[48] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[44] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations all the systems were submitted to an energy minimization protocol for 1,000 steps. The first two nanoseconds of the simulations served to equilibrate systems before the production runs that varied from 40 to 450 ns in the 10 different simulations that were carried out. The equilibration step consisted of 500 ps of simulation where the protein backbone was restrained and 1.5 ns where the system was completely free and no restriction or force was applied. During the equilibration the initial temperature was set to zero and was constantly increased by 1 K every 100 MD steps until the desired temperature (300 K) was reached.

To characterize the coupling between Doc and Coh, we performed SMD simulations[49] of constant velocity stretching (SMD-CV protocol) employing three different pulling speeds: 1.25, 0.625 and 0.25 Å ns^{-1}. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically (center of mass of ASN5), and moving a second restraint point, at the end of the Coh domain (center of mass of GLY210), with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the *z*-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the *z*-axis. Owing to the flexibility of the linkers, this approach reproduces the experimental set-up. All analyses of MD trajectories were carried out employing VMD[42] and its plug-ins. Secondary structures were assigned using the Timeline plug-in, which employs STRIDE criteria[50]. Hydrogen bonds were assigned based on two geometric criteria for every trajectory frame saved: first, distances between acceptor and hydrogen should be <3.5 Å; second, the angle between hydrogen-donor-acceptor should be $<30°$. Surface contact areas of interacting residues were calculated employing Volarea[51] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling spherical probe that is screened around the area of Doc exposed to Coh and also Coh area exposed to Doc.

References

1. Doi, R. H. & Kosugi, A. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* **2,** 541–551 (2004).
2. Carvalho, A. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin—dockerin complex. *Proc. Natl Acad. Sci. USA* **100,** 13809–13814 (2003).
3. Smith, S. P. & Bayer, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* **23,** 686–694 (2013).
4. Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. From cellulosomes to cellulosomics. *Chem. Rec.* **8,** 364–377 (2008).
5. Demain, A. L., Newcomb, M. & Wu, J. H. D. Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* **69,** 124–154 (2005).
6. Jindou, S. *et al.* Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. *FEMS Microbiol. Lett.* **285,** 188–194 (2008).
7. Ding, S. Y. *et al.* Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* **183,** 1945–1953 (2001).
8. Rincon, M. T. *et al.* Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS ONE* **5,** e12476 (2010).
9. Himmel, M. E. *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315,** 804–807 (2007).
10. Fierobe, H.-P. *et al.* Degradation of cellulose substrates by cellulosome chimeras Substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* **277,** 49621–49630 (2002).
11. Valbuena, A. *et al.* On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proc. Natl Acad. Sci. USA* **106,** 13791–13796 (2009).
12. Salama-Alber, O. *et al.* Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288,** 16827–16838 (2013).
13. Adams, J. J., Webb, B. A., Spencer, H. L. & Smith, S. P. Structural characterization of type ii dockerin module from the cellulosome of *Clostridium thermocellum*: calcium-induced effects on conformation and target recognition. *Biochemistry* **44,** 2173–2182 (2005).
14. Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **103,** 305–310 (2006).
15. Sikora, M. & Cieplak, M. Mechanical stability of multidomain proteins and novel mechanical clamps. *Proteins Struct. Funct. Bioinf.* **79,** 1786–1799 (2011).
16. Brunecky, R. *et al.* Structure and function of the Clostridium thermocellum cellobiohydrolase A X1-module repeat: enhancement through stabilization of the CbhA complex. *Acta. Crystallogr.* **68,** 292–299 (2012).
17. Stahl, S. W. *et al.* Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl Acad. Sci. USA* **109,** 20431–20436 (2012).
18. Jobst, M. A., Schoeler, C., Malinowska, K. & Nash, M. A. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.* **82,** e50950 (2013).
19. Otten, M. *et al.* From genes to protein mechanics on a chip. *Nat. Methods* **11,** 1127–1130 (2014).
20. Livadaru, L., Netz, R. R. & Kreuzer, H. J. Stretching response of discrete semiflexible polymers. *Macromolecules* **36,** 3732–3744 (2003).
21. Hugel, T., Rief, M., Seitz, M., Gaub, H. & Netz, R. Highly stretched single polymers: atomic-force-microscope experiments versus *ab-initio* theory. *Phys. Rev. Lett.* **94,** 048301 (2005).
22. Puchner, E. M., Franzen, G., Gautel, M. & Gaub, H. E. Comparing proteins by their unfolding pattern. *Biophys. J.* **95,** 426–434 (2008).
23. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor—ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).
24. Morfill, J. *et al.* Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy. *Biophys. J.* **93,** 3583–3590 (2007).
25. Berkemeier, F. *et al.* Fast-folding α-helices as reversible strain absorbers in the muscle protein myomesin. *Proc. Natl Acad. Sci. USA* **108,** 14139–14144 (2011).
26. Bertz, M., Wilmanns, M. & Rief, M. The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc. Natl Acad. Sci. USA* **106,** 13307–13310 (2009).
27. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402,** 100–103 (1999).
28. Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond? *Science* **283,** 1727–1730 (1999).
29. Xue, Y., Li, X., Li, H. & Zhang, W. Quantifying thiol-gold interactions towards the efficient strength control. *Nat. Commun.* **5,** 4348 (2014).
30. Bomble, Y. J. *et al.* Modeling the self-assembly of the cellulosome enzyme complex. *J. Biol. Chem.* **286,** 5614–5623 (2011).
31. Sotomayor, M. & Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **316,** 1144–1148 (2007).
32. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **271,** 997–999 (1996).
33. Thomas, W. *et al.* Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophys. J.* **90,** 753–764 (2006).
34. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *Biotechniques* **26,** 680–682 (1999).
35. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
36. Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28,** e78 (2000).
37. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.* **41,** 207–234 (2005).
38. Zimmermann, J. L., Nicolaus, T., Neuert, G. & Blank, K. Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments. *Nat. Protoc.* **5,** 975–985 (2010).
39. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1,** 280–285 (2006).
40. Gumpp, H., Stahl, S. W., Strackharn, M., Puchner, E. M. & Gaub, H. E. Ultrastable combined atomic force and total internal fluorescence microscope. *Rev. Sci. Instrum.* **80,** 063704 (2009).
41. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64,** 1868 (1993).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14,** 33–38 (1996).
43. Kalé, L. *et al.* NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151,** 283–312 (1999).
44. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
45. Best, R. B. *et al.* Optimization of the additive CHARMM All-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ 1and χ 2dihedral Angles. *J. Chem. Theory Comput.* **8,** 3257–3273 (2012).
46. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
47. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–934 (1983).
48. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

49. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72,** 1568–1581 (1997).
50. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinf.* **23,** 566–579 (1995).
51. Ribeiro, J. V., Tamames, J. A. C., Cerqueira, N. M. F. S. A., Fernandes, P. A. & Ramos, M. J. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug Des.* **82,** 743–755 (2013).

Acknowledgements

We gratefully acknowledge funding from an advanced grant of the European Research Council (Cellufuel Grant 294438) and from DFG SFB 1032 and the Excellence Cluster Center for Integrated Protein Science Munich. This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and the National Science Foundation (NSF, MCB-1157615 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and the NCSA Blue Waters sustained-petascale supercomputer as part of the general allocations (Simulations of Cellulosomal Subunits: Components of a Molecular Machinery for Depolymerization of Feedstock for Production of Second Generation Biofuels, to K.S.). A grant to E.A.B., H.E.G. and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science - The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland.

Author contributions

Performed and analysed SMFS experiments: C.S., K.H.M., L.F.M., M.A.J., E.D. and M.A.N.; performed and analysed MD simulations: R.C.B. and K.S.; provided proteins and DNA cloning vectors: W.O., D.B.F. and E.A.B.; wrote and edited the manuscript: C.S., K.H.M., R.C.B., E.A.B., K.S., H.E.G. and M.A.N.; supervised research: E.A.B., K.S., H.E.G. and M.A.N.

Additional information

Accession codes: Plasmids used in this study are available through Addgene (https://www.addgene.org) under the following accession codes: Xylanase-Xmodule-Dockerin: 60865; Cohesin-CBM: 60866.

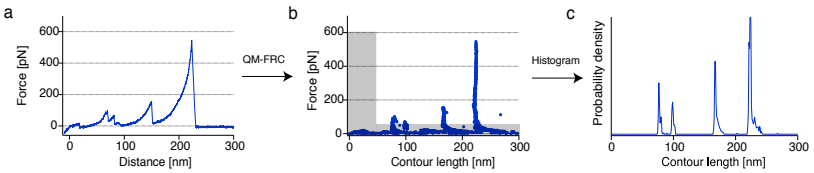
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

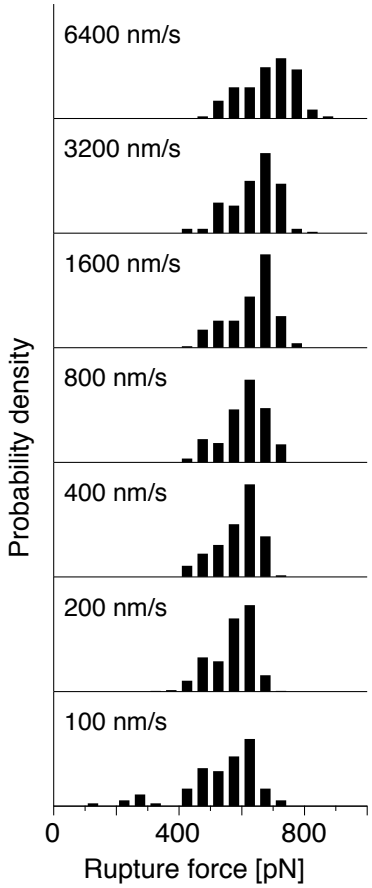
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* 5:5635 doi: 10.1038/ncomms6635 (2014).

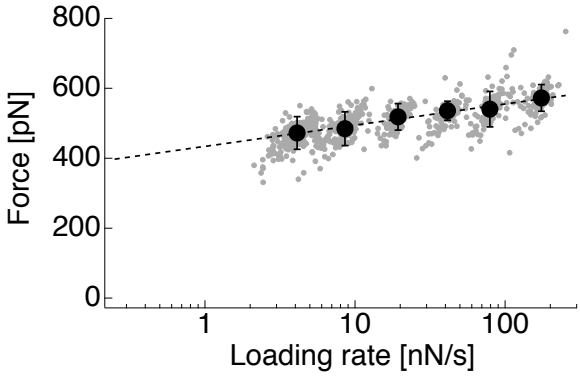
Supplementary Figures



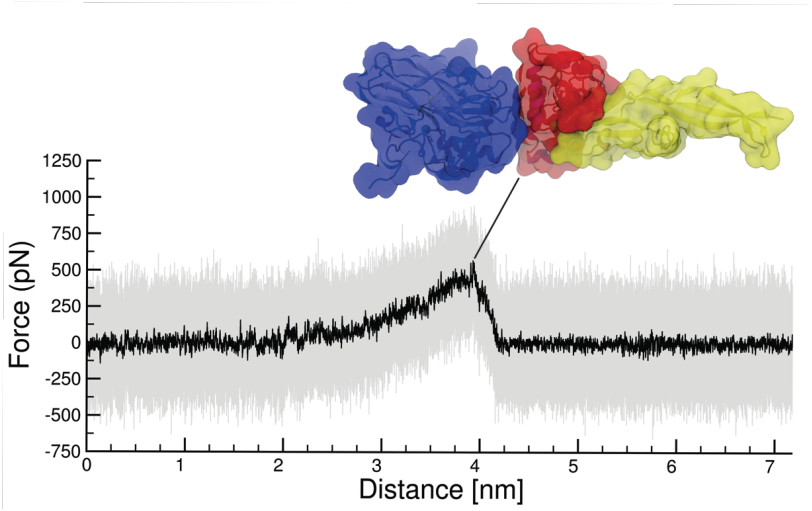
Supplementary Fig. 1: Assembly of contour length histograms. **a** Force-extension traces are transformed into contour length space using a QM-corrected FRC model with parameters $\gamma = 41°$, and $b = 0.11\,\text{nm}$. **b** In force-contour length space, force and contour length thresholds are applied and the data are histogrammed with a bin width of 1 nm to obtain the histogram in **c**. To obtain a master histogram, individual histograms reflecting a specific unfolding pathway are cross-correlated and aligned by offsetting by the maximum correlation value.



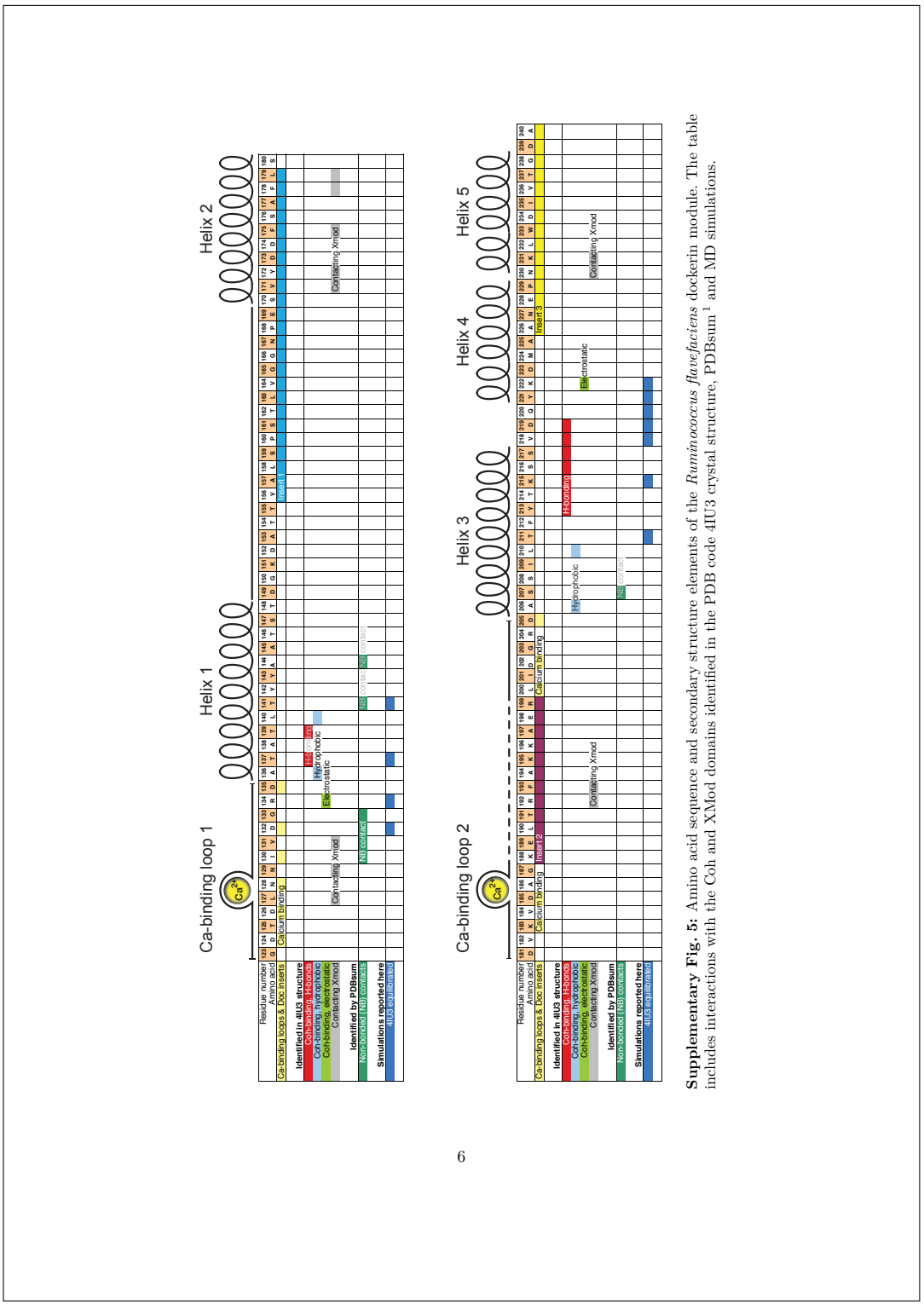
Supplementary Fig. 2: Complex rupture force histograms for pulling speeds ranging from 100 $\mathrm{nm\,s^{-1}}$ to 6400 $\mathrm{nm\,s^{-1}}$. Pulling speeds are indicated next to the histograms. Only traces with an intact XMod were taken into account (no XMod unfolding observed, corresponding to Fig. 2, trace 1). At the slowest pulling speed data suggest the presence of a lower rupture force population.



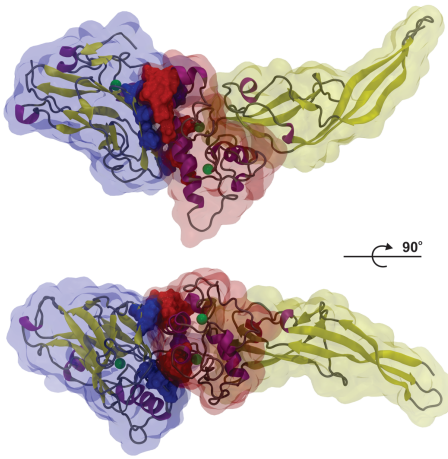
Supplementary Fig. 3: Dynamic force spectrum for XMod unfolding obtained from 654 force-extension traces. The gray points show single XMod unfolding events. Black circles represent the most probable rupture forces and loading rates obtained by Gaussian fitting at each pulling speed. Error bars are ± 1 standard deviation. The dashed line is a least squares fit to the Bell-Evans model that yielded $\Delta x = 0.15\,\mathsf{nm}$ and $k_{off} = 2.6 \times 10^{-6}\,\mathsf{s}^{-1}$.



Supplementary Fig. 4: Force distance trace obtained by SMD at a pulling speed of 0.25 Å ns^{-1}. Force values at each time step are shown in gray, with average force calculated every 200 ps in black. The inset is a snapshot of the XMod-Doc:Coh complex immediately prior to rupture. XMod is shown in yellow, Doc in red and Coh in blue.



Supplementary Fig. 5: Amino acid sequence and secondary structure elements of the *Ruminococcus flavefaciens* dockerin module. The table includes interactions with the Coh and XMod domains identified in the PDB code 4IU3 crystal structure, PDBsum [1] and MD simulations.



Donor	Acceptor	Prevalence
LYS215-S	GLU105-S	87%
ASN107-S	ASP219-S	48%
ASP219-B	ASN107-S	39%
THR141-S	ASP118-S	31%
LYS156-S	ASP132-S	22%

Supplementary Fig. 6: Hydrogen bond contacts between XMod-Doc (yellow and red surface, respectively) and Coh (blue surface). The residues that have hydrogen bonds lasting for more than 10% of the simulation time are represented in a glossy surface. In the bottom of the figure the five most prevalent hydrogen bond interactions are presented. The letter S or B indicate if the respective interaction is made by the amino acid side chain or backbone.

Supplementary Tables

Module	Xylanase	CBM	X-module	Cohesin	Dockerin
No. amino acids, N_A	260 (378)	159	117	205	119
Folded length, L_F [nm]	6	2	7	2	2
Expected increment, ΔL_E [nm]	89	56	36	72	42
Observed increment, [nm]	90 ± 4	55 ± 3	34 ± 2	–	–

Supplementary Table 1: Domain assignment of observed contour length increments. The expected contour length increment (ΔL_E) for each protein domain was calculated according to $\Delta L_E = N_A \cdot 0.365\,\mathrm{nm} - L_F$, where L_F is the folded length, N_A is the number of amino acids, and $0.365\,\mathrm{nm}^2$ is the length per stretched amino acid. L_F was measured for Xyn, CBM, and XDoc:Coh from PDB structures 1R85, 1NBC, and 4IU3, respectively. For the Xyn domain, only amino acids located C-terminal of the C129 mutation which served as attachment point are considered. Errors for the observed increments were determined from Gaussian fits to the combined contour length histogram shown in Fig. 2b.

Supplementary Notes

Supplementary Note 1: QM-FRC Model for Polymer Elasticity

The freely rotating chain model[3] considers bonds of length b, connected by a fixed angle γ. The torsional angles are not restricted. The stretching behavior in the FRC picture is given by

$$\frac{x}{L} = \begin{cases} \frac{Fa}{3k_BT} & \text{for} \quad \frac{Fb}{k_BT} < \frac{b}{p} \\ 1 - \left(\frac{4Fp}{k_BT}\right)^{-\frac{1}{2}} & \text{for} \quad \frac{b}{p} < \frac{Fb}{k_BT} < \frac{p}{b} \\ 1 - \left(\frac{cFb}{k_BT}\right)^{-1} & \text{for} \quad \frac{p}{b} < \frac{Fb}{k_BT} \end{cases} \tag{1}$$

where $a = b\frac{1+\cos\gamma}{(1-\cos\gamma)\cos\frac{\gamma}{2}}$ is the Kuhn length, and $p = b\frac{\cos\frac{\gamma}{2}}{|\ln(\cos\gamma)|}$ is the effective persistence length in the FRC picture.

To account for backbone elasticity of the polypeptide chain at high force, quantum mechanical *ab-initio* calculations can be used to obtain the unloaded contour length at zero force. A polynomial approximation to these calculations can be used to obtain the unloaded contour length at zero force L_0:

$$F = \gamma_1 \left(\frac{L}{L_0} - 1\right) + \gamma_2 \left(\frac{L}{L_0} - 1\right)^2 \tag{2}$$

where the $\gamma_1 = 27.4\,\mathrm{nN}$, and $\gamma_2 = 109.8\,\mathrm{nN}$ are the elastic coefficients reported for polypeptides[4].

Supplementary Note 2: Bell-Evans Model for Mechanically Induced Receptor Ligand Dissociation

The Bell-Evans model was used to estimate the distance to the transition state (Δx) and the natural off-rate (k_{off}) of individual rupture events:

$$\langle F \rangle = \frac{k_BT}{\Delta x} \ln \frac{\Delta x \cdot \dot{F}}{k_{off} k_BT} \tag{3}$$

where k_B is Boltzmann's constant, T is the temperature and $\dot{F}$ is the loading rate at the point of rupture.

Supplementary Methods

Materials

Silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation) with a nominal spring constant of 100 pN/nm (25 kHz resonance frequency in water) were used. Circular coverglasses, 2.4 cm in diameter, were obtained from Menzel Gläser (Braunschweig, Germany). 3-Aminopropyl dimethyl ethoxysilane (APDMES) was purchased from ABCR GmbH (Karlsruhe, Germany). NHS-PEG-Maleimide ($5\,\mathrm{kDa}$) was purchased from Rapp Polymer (Tübingen, Germany). Immobilized TCEP Disulfide Reducing Gel was obtained from Thermo Scientific (Pittsburgh, PA). The following standard chemicals were obtained from Carl Roth (Karlsruhe, Germany) and used as received: tris(hydroxymethyl)aminomethane (TRIS, >99% p.a.), $\mathrm{CaCl_2}$ (>99% p.a.), sodium borate (>99.8% p.a), NaCl (>99.5% p.a.), ethanol (>99% p.a.), and toluene (>99.5% p.a.). Borate buffer was 150 mM, pH 8.5. The measurement buffer for force spectroscopy was Tris-buffered saline (TBS, 25 mM TRIS, 75 mM NaCl, pH 7.2) supplemented with $\mathrm{CaCl_2}$ to a final concentration of 1 mM. All buffers were filtered through a sterile $0.2\,\mu\mathrm{m}$ polyethersulfone membrane filter (Nalgene, Rochester, NY, USA) prior to use.

Protein Sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

Xyn-XModDoc

Xylanase T129C
Linker or extra residues
X-module
Dockerin type III

M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q R Y K N E F T I G A
A V E P Y Q L Q N E K D V Q M L K R H F N S I V A E N V M K P I S I Q P E E G K
F N F E Q A D R I V K F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E
G K P M V N E C D P V K R E Q N K Q L L L K R L E T H I K T I V E R Y K D D I K
Y W D V V N E V V G D D G K L R N S P W Y Q I A G I D Y I K V A F Q A A R K Y G
G D N I K L Y M N D Y N T E V E P K R T A L Y N L V K Q L K E E G V P I D G I G
H Q S H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L D V S M Y G
W P P R A Y P T Y D A I P K Q K F L D Q A A R Y D R L F K L Y E K L S D K I S N
V T F W G I A D N H T W L D S R A D V Y Y D A N G N V V V D P N A P Y A K V E K
G K G K D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T S A V K T Q
Y V E I E S V D G F Y F N T E D K F D T A Q I K K A V L H T V Y N E G Y T G D D
G V A V V L R E Y E S E P V D I T A E L T F G D A T P A N T Y K A V E N K F D Y
E I P V Y Y N N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V D G R
D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T L V G G N P E S V Y D D
F S A F L S D V K V D A G K E L T R F A K K A E R L I D G R D A S S I L T F Y T
K S S V D Q Y K D M A A N E P N K L W D I V T G D A E E E

Coh-CBM C2A, C63S

CBM (C2A, C63S)
Linker or extra residues
CohIII
ybbR-Tag

M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D
A A G Q T V T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G
A Y A K K G A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M
W T V E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A
Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I
K A G E P G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A N
T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I D L S
K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G S N G S Y N G I T S N V K
G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R F A K
N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K E P
G E L K L P R S R H H H H H H G S L E V L F Q G P D S L E F I A S K L A

Supplementary References

[1] Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences* **22**, 488–490 (1997).

[2] Dietz, H.; Rief, M. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences* **103**, 1244–1247 (2006).

[3] Livadaru, L.; Netz, R. R.; Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules* **36**, 3732–3744 (2003).

[4] Hugel, T.; Rief, M.; Seitz, M.; Gaub, H.; Netz, R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-Initio Theory. *Physical Review Letters* **94**, 048301 (2005).

4.3 Associated Publication P3

Mapping Mechanical Force Propagation through Biomolecular Complexes

by

Constantin Schoeler*, Rafael C. Bernardi*, Klara H. Malinowska,

Ellis Durner, Wolfgang Ott, Edward A. Bayer, Klaus Schulten,

Michael A. Nash and Hermann E. Gaub

*(*contributed equally)*

published in

Nano Letters, 15 (11), 7370-7376, (2015)

Reprinted from [14], with permission from American Chemical Society.

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Mapping Mechanical Force Propagation through Biomolecular Complexes

Constantin Schoeler,[†,#] Rafael C. Bernardi,[‡,#] Klara H. Malinowska,[†] Ellis Durner,[†] Wolfgang Ott,[†,§] Edward A. Bayer,[∥] Klaus Schulten,[‡,⊥] Michael A. Nash,[*,†] and Hermann E. Gaub[†]

[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

[‡]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

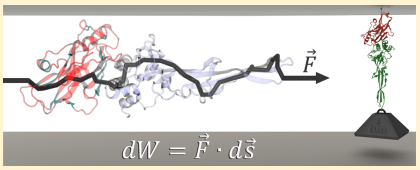
[§]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany

[∥]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

[⊥]Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Here we employ single-molecule force spectroscopy with an atomic force microscope (AFM) and steered molecular dynamics (SMD) simulations to reveal force propagation pathways through a mechanically ultrastable multidomain cellulosome protein complex. We demonstrate a new combination of network-based correlation analysis supported by AFM directional pulling experiments, which allowed us to visualize stiff paths through the protein complex along which force is transmitted. The results implicate specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces.



KEYWORDS: *Force propagation, single molecule force spectroscopy, steered molecular dynamics, network analysis, cohesin−dockerin*

Mechanical forces play a fundamental role in biological systems. Cells are able to sense and respond to mechanical cues in their environment by, for example, modulating gene expression patterns,[1] reshaping the extracellular matrix,[2] or exhibiting differential biochemical activities.[3] At the molecular level, these behaviors are governed by mechanically active proteins. Such proteins are able to sense and respond to force by undergoing conformational changes,[4] exposing cryptic binding sequences,[5] acting synergistically with ion channels,[6] or modulating their function in a variety of ways.[7−9]

Experimental methods including AFM single-molecule force spectroscopy (SMFS) allow direct measurement of molecular mechanical properties. These studies have demonstrated the importance of the shear topology involving parallel breakage of hydrogen bonds in providing mechanical stability to protein folds.[10,11] Many globular domains and protein complexes also exhibit a directional dependence in unfolding mechanics, consisting of stiff and soft axes.[12−18] Pulling geometry can be defined by controlling the positions of the chemical linkages between protein monomer units through a variety of bioconjugate techniques.

Primary sequences of mechanically active proteins are extremely diverse, essentially rendering them undetectable by conventional bioinformatics approaches. Yet, another computational approach, namely, molecular dynamics (MD), allows sampling of structural conformations of large and frequently mechanostable protein complexes.[19,20] Analysis of these conformations from MD trajectories have recently led to the development of network-based correlation methods for investigating signal transmission and allosteric regulation in proteins.[21−23] In network models, local correlations of positional fluctuations in a protein are represented as a web of inter-residue connections. Within such a network, the behavior of nodes that are highly correlated and within close physical proximity can be analyzed to obtain the shortest path between two network nodes (i.e., amino acids). This analysis helps to identify which connecting residues are most important for intramolecular communication.[23−25] Examination of multiple pathways, also known as suboptimal paths, within an acceptable deviation from the optimal path helps to detect the web of nodes critical for transmission of information.

Among MD methods, steered molecular dynamics (SMD) simulations in which external forces are used to explore the response and function of proteins have become a powerful tool especially when combined with SMFS.[6] SMD has been successfully employed in a wide range of biological systems, from the investigation of protein mechanotransduction,[5,26] to

Received: July 9, 2015
Revised: August 5, 2015
Published: August 11, 2015

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Nano Letters Letter

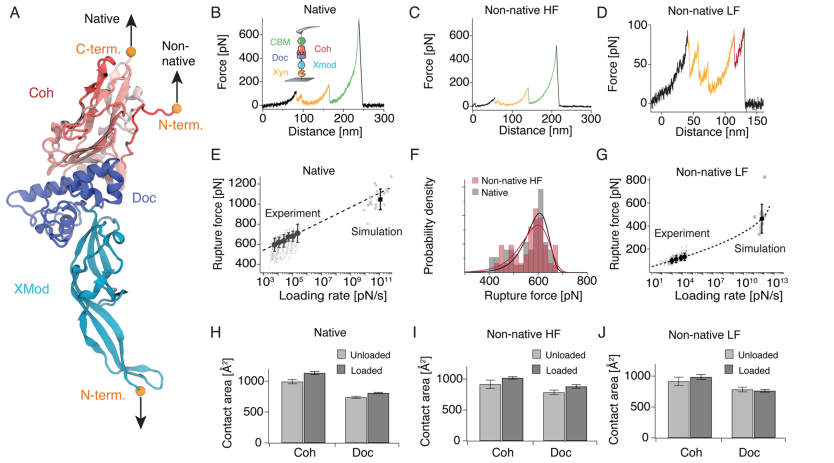


Figure 1. Single molecule force spectroscopy and steered molecular dynamics of XMod-Doc:Coh in two pulling configurations. (A) Crystal structure of the XMod-Doc:Coh complex (PDB 4IU3) with orange spheres marking the termini where force was applied. (B) Experimental unfolding trace for the native pulling configuration at a pulling speed of 1600 nm s^{-1}. The inset shows a schematic of the pulling geometry. Unfolding signatures of the Xyn and CBM marker domains are marked in orange and green, respectively. (C) Experimental unfolding trace for the non-native high force class obtained at a pulling speed of 700 nm s^{-1}. (D) Experimental unfolding trace for the non-native low force class obtained at a pulling speed of 700 nm s^{-1}. The additional 17−19 nm contour length increment attributed to N-terminal Coh unfolding is shown in red. (E) Dynamic force spectrum for XMod-Doc:Coh unbinding in the native geometry obtained from experiment and simulations. Gray points and squares represent the rupture force/loading rate pairs obtained from experiment and simulation, respectively. Black circles represent the most probable rupture force/loading rate obtained from Gaussian fits to the experimental data at six pulling speeds. The black square shows the mean rupture force and loading rate for the simulated rupture events. (F) Rupture force histograms obtained at a pulling speed of 800 nm s^{-1} for the native (gray, $n = 46$) and non-native high force class (red, $n = 48$). Fitted probability densities $p(F)$ are shown as solid black and red lines. Data for both pulling configurations were obtained with the same cantilever to minimize calibration errors. (G) Dynamic force spectrum for XMod-Doc:Coh unbinding in the non-native low force class obtained from experiments and simulation. The same representation as in (E) is used. (H,I,J) Unloaded and loaded surface contact areas for the different pulling geometries ((H) native, (I) non-native high force class, and (J) non-native low force class).

permeability of membrane channels,[27,28] and the characterization of protein−receptor interactions.[29] SMD simulations have also been used to study force propagation through proteins by employing force distribution analysis (FDA).[30,31] In FDA, all pairwise forces, which are usually calculated in MD simulations, are stored in $N \times N$ matrices, where N is the number of atoms.[32] These pairwise forces can then be used to assess a protein's response to a mechanical or allosteric signal.[33] In the FDA approach, atoms under mechanical strain are identified by subtracting forces of both loaded and unloaded states for each pair of interacting atoms.[31] However, to achieve a sufficient signal-to-noise ratio, FDA will often require exhaustive sampling of the conformational space.[32,34] FDA, therefore, requires more computational resources than usual SMD studies, which are frequently already computationally demanding. There is therefore a clear need for new analysis methods that enable visualization of force propagation pathways from a single SMD trajectory.

Here we implemented a novel combination of SMD, network-based correlation analysis, and thermodynamic fluctuation theory, supported by AFM-SMFS experiments to study force propagation through a protein complex subjected to different pulling geometries. We chose an ultrastable receptor−ligand interaction as a model system because of its remarkably high mechanical stability,[29] which effectively improves the signal-to-noise ratio. This complex consists of two interacting protein domains called cohesin (Coh) and dockerin (Doc) that maintain bacterial adhesion of *Ruminococcus flavefaciens* to cellulosic substrates. Doc is found within the same polypeptide chain as a stabilizing ancillary domain called X-module (XMod), located N-terminally of Doc. Based on its position with the *R. flavefaciens* cellulosomal network, Coh is mechanically anchored *in vivo* at its C-terminal end to the cell surface. Our prior work demonstrated that, when force is applied to the complex in the native configuration (i.e., C-terminal Coh, N-terminal XMod-Doc anchor points), the complex is extremely stable, exhibiting high rupture forces of 600−750 pN at loading rates from 1−100 nN s^{-1}.[29] Since the bulk equilibrium affinity of the complex is an unremarkable 20 nM,[35] we hypothesized that the high mechanostability is explained by a catch bond mechanism. AFM rupture force data and SMD simulations supported this prediction, where it was observed that the contact surface area of the two proteins increased as mechanical force was applied.

To characterize the mechanisms behind Coh:Doc high stability, here we additionally pulled the complex apart in a non-native configuration (i.e., N-terminal Coh, N-terminal XMod-Doc anchor points). In the non-native pulling

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

configuration, we found that the complex dissociated along two competing pathways with very different mechanical characteristics.

Our new dynamic network analysis protocol reveals how different mechanical behaviors are attributable to differences in the direction of force transmission across the binding interface. Together, the experiments and simulations depict a simple physical mechanism for achieving high complex rupture forces: the complex directs force along pathways orthogonal to the pulling axis.

Single-Molecule Pulling Experiments and SMD. For SMFS experiments, XMod-Doc was produced as a fusion protein with an N-terminal Xylanase (Xyn) domain. Coh was produced as either an N- or C-terminal fusion domain with a carbohydrate binding module (CBM). These fusion domains were used for site specific immobilization to a glass surface and AFM cantilever to achieve the two loading configurations shown in Figure 1A and further served as marker domains with known unfolding length increments to validate single-molecule interactions and sort SMFS data traces.[36]

For the native pulling configuration found *in vivo*, Coh-CBM and XMod-Doc are loaded from their C- and N-termini, respectively (Figure 1A). A representative unbinding trace for the native pulling configuration is shown in Figure 1B. We measured the loading rate dependence of complex rupture using both experimental and SMD data sets (unbinding trace from SMD shown in Figure 3A) and plotted them on a combined dynamic force spectrum (Figure 1E). The linear Bell model produced fit parameters for the effective distance to the transition state Δx = 0.13 nm, and the zero-force off rate k_{off} = 4.7×10^{-4} s^{-1}. Both experimental and simulation data are well described by a single Bell expression, despite the differences in loading rates between experiments and simulation. The observation suggests that the application of force does not significantly change Δx for this particular configuration.

To test the influence of pulling geometry on mechanical stability, we performed SMFS and SMD on the system where Coh was pulled from the opposite terminus (i.e., non-native N-terminus, cf. Figure 1A). Unlike the native pulling geometry, this geometry exhibited two clearly distinct unbinding pathways that are characterized by different force ranges (high or low) at which the complex dissociated. We refer to these pathways as non-native high force (HF) (Figure 1C) and non-native low force (LF) (Figure 1D).

AFM data traces classified as non-native HF showed similar characteristics as those in the native pulling configuration (cf. Figure 1B,C,F). The non-native LF traces, however, exhibited a markedly different unfolding behavior (Figure 1D). Xyn unfolding (highlighted in orange) was regularly observed, but CBM unfolding was only very rarely observed. The complex usually did not withstand forces high enough to unfold CBM when rupturing along the non-native LF path. Among non-native LF curves, we regularly found an additional contour length increment of 17−19 nm consistent with unfolding of ∼60 amino acids located at the N-terminus of Coh. This unfolding occurred immediately following Xyn unfolding (Figure 1D, red), or alternatively prior to Xyn unfolding, or with a substep (Supplementary Figure S1). Taken together, it appears that partial Coh unfolding from the N-terminus destabilizes the complex, causing lower rupture forces (Figure 1G).

The experimental rupture forces from the non-native HF class were indistinguishable from those arising in the native configuration. To confirm this, we performed additional measurements where both Coh configurations were alternately probed with the same Xyn-XMod-Doc functionalized cantilever (Supplementary Figure S2), eliminating inaccuracies introduced through multiple cantilever calibration. Most probable rupture forces at a pulling speed of 800 nm s^{-1} of 606 and 597 pN for the native configuration and non-native HF class, respectively, were determined in the Bell Evans model (Figure 1F, Supplementary eq S2), demonstrating that the native and non-native HF classes are experimentally indistinguishable.

For the LF class, we analyzed the final complex rupture event and plotted the combined dynamic force spectrum (Figure 1G). Here, simulated and experimentally observed data were not well described by a single Bell expression. In such cases nonlinear models have been developed to obtain kinetic and energetic information from dynamic force spectra.[37,38] To fit the combined data, we used the nonlinear Dudko−Hummer−Szabo (DHS) model (Supplementary eq S3) and obtained values of Δx = 0.42 nm and k_{off} = 0.005 s^{-1}. The DHS model further provides the free energy difference ΔG between the bound state and the transition state as a fit parameter, which was found to be ΔG = 129 k_BT. The model fit produced a distance to transition that was much longer than observed for the native configuration. Independent SMD simulations for the non-native pulling configuration were found to also lead to HF and LF unbinding scenarios (see below, Figure 4A,D, respectively).

The differential solvent contact area was calculated from SMD simulations to estimate the intermolecular contact area in the Doc:Coh complex. In the native configuration, the simulated Doc:Coh contact area increased by 14% and 9% for Coh and Doc, respectively (Figure 1H). For the non-native HF class, the contact area increased by 11% and 12% for Coh and Doc, respectively (Figure 1I). In the non-native LF class, the contact area increased by only 7% for Coh and decreased by 3% for Doc (Figure 1J). Evidently, an increased surface contact area for Doc in the native and non-native HF pathways correlated with high mechanostability of the system.

Force Propagation Theory: A Simple Model. To further understand the observed unbinding pathways, we sought to identify paths through the molecule along which the externally applied load propagates. From thermodynamic fluctuation theory,[39,40] it is known that the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i are related through

$$\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{1}$$

where $\Delta \mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$ and $\mathbf{r}_i$ is the position of atom i. The derivative on the right-hand side of eq 1 states that neighboring atoms i and j will move with high correlation due to an external force $\mathbf{F}_i$ acting on atom i if the coupling between them is strong. Hence, a given element of a correlation matrix $M_{ij} = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle$ will be large in the case of a strong interaction potential between i and j. When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation, while stiff degrees become more important for the dynamics of the system.

Consequently, paths with high correlation of motion describe the paths along which force propagates through the system. To illustrate this behavior for a toy system, we employed the NAMD[41] SMD[42] constant velocity protocol to a test pattern of identical spheres connected with harmonic springs of different

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

stiffness (Figure 2A). The position of one sphere was fixed during the simulation, while another sphere on the opposite side of the structure was withdrawn at constant velocity. The strained structure at the end of the simulation is shown in Figure 2B. We assigned weights to the lines between spheres according to the Pearson correlation coefficient C_{ij} (Supplementary eq S4) between those network nodes (Figure 2C). The Pearson correlation coefficient differs from the left-hand side of eq 1 by a normalization factor $\langle\langle \Delta \mathbf{r}_i^2(t)\rangle\langle \Delta \mathbf{r}_j^2(t)\rangle\rangle^{-1/2}$ and was chosen to make our analysis mathematically more tractable. For a detailed discussion on this choice of correlation measure, see Supporting Information. In a harmonic potential approximation, the equipartition theorem can be applied to this normalization factor resulting in the following expression for C_{ij}:

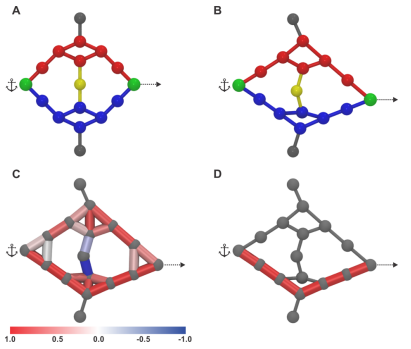


Figure 2. Network analysis test simulation. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red, k; blue, $5k$; yellow, $7.5k$; black, $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs were introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red.

$$C_{ij} = \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i}\sqrt{k_{i,\mathrm{eff}}k_{j,\mathrm{eff}}} \tag{2}$$

where $k_{i,\mathrm{eff}} = \left(\frac{1}{k_{x_i}} + \frac{1}{k_{y_i}} + \frac{1}{k_{z_i}}\right)^{-1}$ and k_{x_i} is the curvature of the potential on atom i in the x direction. For a full derivation, see Supporting Information. Equation 2 illustrates how Pearson correlation is a suitable measure to identify the stiff paths in our simple model. We then used dynamical network analysis implemented in VMD[49] to find the path of highest correlation (Figure 2D). As expected from eq 1, we found this path to be the one connected by the stiff springs.

Force Propagation through XMod-Doc:Coh Complex. The simple pattern of spheres validated our general approach of using local correlations to identify load-bearing pathways through networks. We next employed dynamical network analysis to understand force propagation through the XMod-Doc:Coh complex.

The dynamic networks for the native configuration (unloaded and loaded) are shown in Figure 3B,C, respectively. While the network shows multiple suboptimal paths in the unloaded scenario, the loaded case exhibits a well-defined main path along which force propagates through the system. Interestingly, in the loaded configuration, force propagates through both binding helices of Doc, which results in a force path with large normal components to the unbinding axis close to the binding interface as illustrated in Figure 3D. It had been shown for another ultrastable protein, namely, silk crystalline units, that curving force paths distribute tension through the entire system.[31] A strategy that assumes an indirect path would therefore allow the system to have more time to absorb the tension from the applied force. The result here supports the view that directing the force along a path with significant perpendicular components to the pulling axis leads to high mechanical stability. In a simple mechanical picture, a certain amount of mechanical work, namely $dW = \mathbf{F}\cdot d\mathbf{s}$, is required to separate the two binding interfaces by a distance Δz and break the interaction. In this simplified picture, $d\mathbf{s}$ points along the unbinding axis, whereas the force $\mathbf{F}$ is locally largely perpendicular to this direction. Consequently, a larger force is required to break the interaction than in a scenario where the force path would point along the unbinding axis.

To validate this picture, we repeated the same analysis for the non-native HF and non-native LF pathways. The HF simulation (Figure 4A) exhibited only a small stretching of the flexible N-terminal region of Coh and complex dissociation at approximately 800 pN and a pulling distance around 10 nm. However, the LF case shown in Figure 4D exhibited a stepwise N-terminal Coh unfolding, dissociating at a force of about 480 pN at a pulling distance of about 25 nm. This behavior confirmed our assignment of the experimentally observed 17–19 nm contour length increment to Coh unfolding up to residue 62 in PDB 4IU3.

While the experimental data did not show a detectable difference between the native configuration and the non-native HF class, the propagation of force takes place along a different pathway (Figure 4B). For N-terminal Coh pulling, helix 3 of Doc is not involved in the propagation of force as it is for the native geometry. In the native configuration, force propagates through the center of Coh, while for non-native HF the path is shifted toward the side of the molecule. Despite these differences, there is a common feature between the native and non-native HF pathways. At the binding interface, the pathway again shows pronounced components perpendicular to the unbinding axis (cf. Figure 4C), suggesting that this feature is indeed responsible for the exceptional mechanical strength observed for these two unbinding pathways.

Figure 4E shows the force propagation pathway for the non-native LF class prior to rupture. Due to the unfolding of the N-terminal Coh segment, the propagation of force is shifted even further away from the central portion of Coh than for the non-native HF class. Interestingly, force is propagated through the small helical segment of Coh (ALA167-GLN179), a portion of the molecule that is not involved in force propagation for any of the other analyzed trajectories. Unlike in the aforementioned scenarios, there is no pronounced tendency for perpendicular force components at the binding interface for the non-native LF class. In fact, the force is propagated along a path largely parallel to the pulling axis (cf. Figure 4F). In cases where force

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370–7376

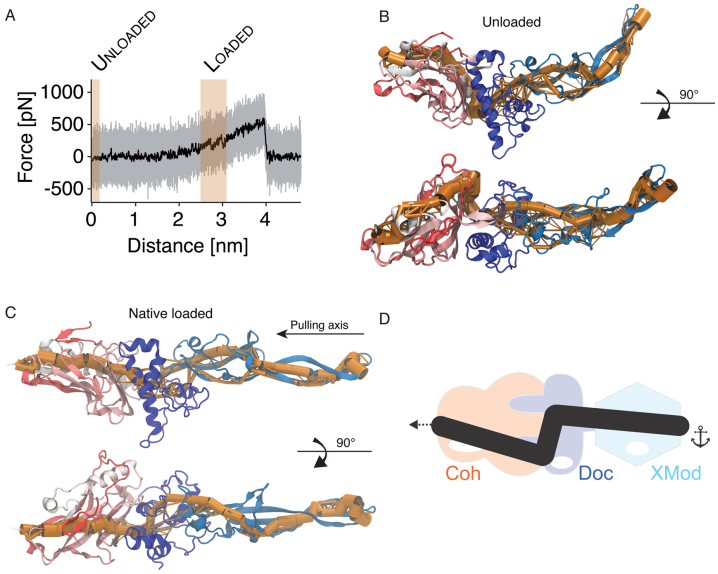


Figure 3. Force propagation through XMod-Doc:Coh in the native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray. The black line represents a moving average with a box size of 500 steps. The highlighted red areas denote the windows where dynamic networks and contact areas were calculated. (B) Network paths for the unloaded system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (C) Network paths for the loaded system. A detailed 2D representation of the pathway, highlighting the amino acids present in the pathway, is shown in Supplementary Figure S5. (D) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis.

propagation occurs parallel to the pulling axis, as in Figure 4E, low mechanical stability was observed.

The aforementioned force propagation architecture along with the effect of increasing contact surface area upon mechanical loading combine for elevated mechanostability of the system.[29] In cases where we observed an N-terminal Coh unfolding of 62 amino acids in the non-native geometry, the system was no longer able to summon this mechanism, causing dissociation at much lower forces.

Previously, our groups have reported on a family of mechanically stable protein ligand receptor complexes that are key building blocks of cellulosomes,[29,44−46] the multienzyme complexes used by select anaerobic bacteria to digest lignocellulose. However, the molecular origins of the stability of these complexes remained largely unclear. An initial clue was obtained when, in a previous work, we were able to show that contact surface area of the two proteins increased as mechanical force was applied.[29] In a different study,[47] coarse-grained MD simulations showed much smaller rupture forces at similar loading rates both for native and non-native pulling than we report here. This disagreement is likely due to the inability of the coarse-grained model to capture the rearrangement of amino acid side chains observed here. As we demonstrated, force propagation calculation from network-based correlation analysis helped in investigating the dramatic effect on the mechanical stability of the Doc:Coh interaction when different pulling geometries are applied. Our methodological approach, to the best of our knowledge, has never been applied even though network analysis of SMD trajectories was performed before to probe the mechanism of allosteric regulation in imidazole glycerol phosphate synthase.[48]

In summary, for both unbinding cases where we observed high mechanostability, we found that across the binding interface, force propagated along paths with strong normal components to the pulling direction. Such a behavior was not observed for the non-native LF class, where, presumably due to N-terminal Coh unfolding, the system was no longer able to direct the force across the binding interface at high angles. From these findings, we conclude that the ultrastable complex formed by Coh and Doc achieves its remarkable mechanostability by actively directing an externally applied force toward an unfavorable angle of attack at the binding interface, consequently requiring more force to achieve a given amount of separation along the pulling direction. Our results show that this mechanically stable complex uses an architecture that exploits simple geometrical and physical concepts from Newtonian mechanics to achieve high stability against external forces. The analytical framework derived here provides a basis for developing a deeper understanding of the functioning of various mechanoactive proteins that are crucial for physiolog-

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

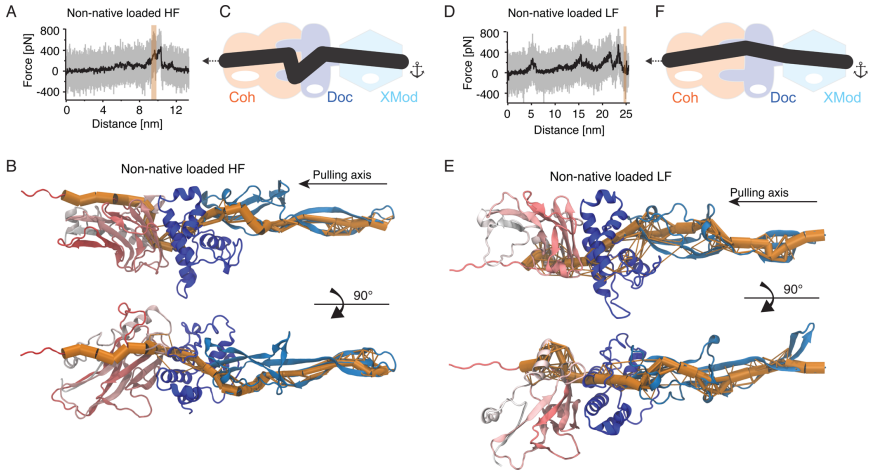


Figure 4. Force propagation through XMod-Doc:Coh in the non-native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500 steps. Note that this computational pulling experiment revealed a high-force behavior. (B) Network pathways calculated from dynamical network analysis for the non-native HF trajectory. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S6. (C) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis. (D) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500. This computational pulling experiment revealed partial Coh unfolding that led to LF behavior. (E) Network pathways for the non-native LF scenario. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S7. (F) Schematic model of force propagation across the Coh:Doc binding interface. Unlike in both HF scenarios, force propagates across the binding interface mostly along the unbinding axis.

ically relevant processes such as mechanotransduction, cellular mechanosensing, and pathogenesis. Additionally, it could provide a design platform for development of artificial mechanoactive systems with applications as tissue engineering scaffolds or components in engineered nanomaterials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b02727.

Details of experimental materials and methods, steered molecular dynamics simulations, and additional theoretical discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.nash@lmu.de.

Author Contributions

#These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Zaida Luthey-Schulten, Helmut Grubmüller, Marcelo Melo, and Jonathan Lai for helpful discussions. The authors gratefully acknowledge funding from an advanced grant to HEG from the European Research Council (Cellufuel Grant 294438). This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and from the Energy Biosciences Institute (UCB BP 2014OOO4J01 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and NERSC/Edison supercomputer as part of the DoE ALCC program. Resources of the National Energy Research Scientific Computing Center (NERSC) are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A grant to E.A.B., H.E.G., and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. The authors appreciate the support of the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nano-systems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science—The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland, and from the LMU Excellence Initiative.

REFERENCES

(1) Bao, G.; Suresh, S. *Nat. Mater.* **2003**, *2*, 715−725.
(2) Grodzinsky, A. J.; Levenston, M. E.; Jin, M.; Frank, E. H. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 691−713.
(3) Kumar, S.; Weaver, V. M. *Cancer Metastasis Rev.* **2009**, *28*, 113−127.
(4) Vogel, V.; Thomas, W. E.; Craig, D. W.; Krammer, A.; Baneyx, G. *Trends Biotechnol.* **2001**, *19*, 416−423.
(5) del Rio, A.; Perez-Jimenez, R.; Liu, R.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P. *Science* **2009**, *323*, 638−641.
(6) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−1148.
(7) Engel, A.; Muller, D. J. *Nat. Struct. Biol.* **2000**, *7*, 715−718.
(8) Noy, A.; Friddle, R. W. *Methods* **2013**, *60*, 142−150.
(9) De Yoreo, J. J.; Chung, S.; Friddle, R. W. *Adv. Funct. Mater.* **2013**, *23*, 2525−2538.
(10) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys. J.* **1998**, *75*, 3008−3014.
(11) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−1112.
(12) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. *Biophys. J.* **2012**, *103*, 2361−2368.
(13) Lee, W.; Zeng, X.; Rotolo, K.; Yang, M.; Schofield, C. J.; Bennett, V.; Yang, W.; Marszalek, P. E. *Biophys. J.* **2012**, *102*, 1118−1126.
(14) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820−17825.
(15) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12724−12728.
(16) Carrión-Vázquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738−743.
(17) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731−737.
(18) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(19) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64−74.
(20) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(21) Ribeiro, A. S. T.; Ortiz, V. *J. Phys. Chem. B* **2015**, *119*, 1835−1846.
(22) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.
(23) Van Wart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.
(24) Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z. *FEBS Lett.* **2010**, *584*, 376−386.
(25) Chennubhotla, C.; Bahar, I. *Mol. Syst. Biol.* **2006**, *2*, 1−13.
(26) Hytönen, V. P.; Vogel, V. *PLoS Comput. Biol.* **2008**, *4*, e24.
(27) Khalili-Araghi, F.; Gumbart, J.; Wen, P.-C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2009**, *19*, 128−37.
(28) Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 271−281.
(29) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; Gaub, H. E.; Nash, M. A. *Nat. Commun.* **2014**, *5*, 5635.
(30) Stacklies, W.; Vega, M. C.; Wilmanns, M.; Gräter, F. *PLoS Comput. Biol.* **2009**, *5*, e1000306.
(31) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. *Biophys. J.* **2009**, *96*, 3997−4005.
(32) Stacklies, W.; Seifert, C.; Graeter, F. *BMC Bioinf.* **2011**, *12*, 101.
(33) Seifert, C.; Gräter, F. *Biophys. J.* **2012**, *103*, 2195−2202.
(34) Palmai, Z.; Seifert, C.; Gräter, F.; Balog, E. *PLoS Comput. Biol.* **2014**, *10*, e1003444.
(35) Salama-Alber, O.; Jobby, M. K.; Chitayat, S.; Smith, S. P.; White, B. A.; Shimon, L. J. W.; Lamed, R.; Frolow, F.; Bayer, E. A. *J. Biol. Chem.* **2013**, *288*, 16827−16838.
(36) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. *Biophys. J.* **2008**, *95*, 426−434.
(37) Friddle, R. W.; Noy, A.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13573−13578.
(38) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(39) Greene, R. F.; Callen, H. B. *Phys. Rev.* **1951**, *83*, 1231−1235.
(40) Erman, B. *Phys. Biol.* **2011**, *8*, 056003.
(41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(42) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(43) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(44) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127−1130.
(45) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.
(46) Jobst, M. A.; Schoeler, C.; Malinowska, K. H.; Nash, M. A. *J. Visualized Exp.* **2013**, *82*, e50950.
(47) Wojciechowski, M.; Thompson, D.; Cieplak, M. *J. Chem. Phys.* **2014**, *141*, 245103.
(48) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. *Biochemistry* **2007**, *46*, 2156−2173.
(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Supporting Information

Mapping mechanical force propagation through biomolecular complexes

Constantin Schoeler[a], Rafael C. Bernardi[b], Klara H. Malinowska[a], Ellis Durner[a], Wolfgang Ott[a,c], Edward A. Bayer[d], Klaus Schulten[b,e], Michael A. Nash[*a], and Hermann E. Gaub[a]

[a]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
[c]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany
[d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
[e]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[*] michael.nash@lmu.de

1 Materials and Methods

1.1 Site Directed Mutagenesis

We performed site-directed mutagenesis of *Ruminococcus flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus*[1] were subjected to QuikChange mutagenesis to install the mutations described in the prior paper[2]. All mutagenesis products were confirmed by DNA sequencing analysis.

1.2 Expression and Purification of Cysteine-Mutated Xyn-XMod-Doc

The Xyn(T129C)-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16°C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20 mM imidazole and 2mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialyzed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at ~ 20°C. The concentrations of the protein stock solutions were determined to be ~ 5 mg/mL by absorption spectrophotometry.

1.3 Expression and Purification of Coh-CBM and mutated Coh-CBM C63S

The Coh-CBM (C63S) fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicole containing ZYM-5052 media[3] overnight at 22°C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at -80°C. The concentrations of the protein stock solutions were determined to be ~ 5 mg/mL by absorption spectrophotometry.

1.4 Sample Preparation

Cantilevers and cover glasses were functionalized according to previously published protocols[4]. Briefly, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyl-XMod-Doc proteins were reduced for 2 hours at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with TBS measurement buffer, centrifuged at 850 rcf for 3 minutes, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with TBS to remove

unbound proteins, and stored under TBS prior to force spectroscopy measurements. Site specific immobilization of the Coh-CBM-ybbR fusion proteins to PEGylated cantilevers or coverglasses was carried out according to previously published protocols[5]. Briefly, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer for 1h at room temperature. Levers or surfaces were then rinsed with TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 hours at room 37°. Finally, surfaces or levers were subjected to a final rinse with TBS and stored under TBS prior to measurement.

1.5 Single Molecule Force Spectroscopy Measurements

SMFS measurements were performed on a custom built AFM controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise / equipartition method. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2–6.4 μm/s. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18-24 hours. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw datasets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single molecule interactions with CBM and Xyn unfolding length increments. We sorted the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: the xylanase (∼89 nm) and the CBM (∼56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

1.6 Data Analysis

Data were analyzed using slight modifications to previously published protocols[4;6;7]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11\,\mathrm{nm}$ connected by a fixed angle $\gamma = 41°$ and and assembled into barrier position histograms using cross-correlation. For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs. time trace in the immediate vicinity prior to the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying probability density fits to histograms of rupture forces and loading rates at each pulling speed.

1.7 Molecular Dynamics Simulations

Connecting dynamics to structural data from diverse experimental sources, molecular dynamics simulations allow one to explore off-equilibrium properties of protein structure complexes in unparalleled detail[8]. More specifically, molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes[9]. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97Å resolution and is available at the protein data bank (PDB:4IU3). The system was then solvated and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter-ions,

which were randomly arranged in the solvent. Total system size was approximately 580k atoms. The MD simulations in the present study were performed employing the molecular dynamics package NAMD[10;11]. The CHARMM36 force field[12;13] along with the TIP3 water model[14] was used to describe all systems. The simulations were carried out assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME)[15] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[11] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. The first two nanoseconds of the simulations served to equilibrate systems before the production runs, which varied from 200 ns to 1.3 μs in the different simulations. To characterize the coupling between dockerin and cohesin, we performed SMD simulations[16] of constant velocity stretching (SMD-CV protocol) with pulling speed of 0.25 Å/ns. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically, and moving a second restraint point, at the end of the Coh domain, with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic pulling spring is then monitored during the time of the molecular dynamics simulation. All analyses of MD trajectories were carried out employing VMD[17] and its plugins. Surface contact areas of interacting residues were calculated employing Volarea[18] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling sphere that is scanned around the area of the dockerin exposed to the cohesin and also the cohesin area exposed to the dockerin. The Network View plugin[19] on VMD[17] was employed to perform dynamical network analysis. A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.[20], nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as $w_{ij} = -log\,(\mid C_{ij} \mid)$, where C_{ij} is the correlation matrix calculated with Carma[21]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was $-log\,(0.5) = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently.

2 Protein Sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

2.1 HIS-Xyn(T128C)-XDoc

X-module
Dockerin type III

Xylanase
Linker or extra residues

M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q R Y K N E F T I G A
A V E P Y Q L Q N E K D V Q M L K R H F N S I V A E N V M K P I S I Q P E E G K
F N F E Q A D R I V K F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E
G K P M V N E C D P V K R E Q N K Q L L L K R L E T H I K T I V E R Y K D D I K
Y W D V V N E V V G D D G K L R N S P W Y Q I A G I D Y I K V A F Q A A R K Y G
G D N I K L Y M N D Y N T E V E P K R T A L Y N L V K Q L K E E G V P I D G I G
H Q S H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L D V S M Y G
W P P R A Y P T Y D A I P K Q K F L D Q A A R Y D R L F K L Y E K L S D K I S N
V T F W G I A D N H T W L D S R A D V Y Y D A N G N V V V D P N A P Y A K V E K
G K G K D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T S A V K T Q
Y V E I E S V D G F Y F N T E D K F D T A Q I K K A V L H T V Y N E G Y T G D D
G V A V V L R E Y E S E P V D I T A E L T F G D A T P A N T Y K A V E N K F D Y
E I P V Y Y N N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V D G R
D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T L V G G N P E S V Y D D
F S A F L S D V K V D A G K E L T R F A K K A E R L I D G R D A S S I L T F Y T
K S S V D Q Y K D M A A N E P N K L W D I V T G D A E E E

2.2 Coh-CBM(C2A,C63S)-HIS-ybbR

CohIII
CBM (C2A, C63S)
ybbR-Tag
Linker or extra residues

M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D
A A G Q T V T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G
A Y A K K G A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M
W T V E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A
Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I
K A G E P G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A N
T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I D L S
K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G S N G S Y N G I T S N V K
G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R F A K
N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K E P
G E L K L P R S R H H H H H H G S L E V L F Q G P D S L E F I A S K L A

2.3 CBM(T2C)-Coh-HIS

CBM (T2C)
CohIII
Linker or extra residues

M C N T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I
D L S K L T L R Y Y Y T V D G Q K D Q T F W C D H A A I I G S N G S Y N G I T S

N V K G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R
F A K N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G
K E P G G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A L T
D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D A A G Q T V
T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G A Y A K K G
A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M W T V E L K
V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A Q G K L M Q
A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I K A G E P L
E H H H H H H

3 Supplementary Discussion

The Pearson correlation matrices of the Xmod-Doc:Coh complex before and after applying force in the native pulling configuration are presented in Supplementary Figure S3 and S4, respectively. For the unloaded complex, movements within Doc domain are seen to be highly correlated, while XMod is seen to be divided into two anti-correlated sub-domains, one comprising the β-sheet fragment close to the N-terminus (residues 5-15 and 45-66) and the other constituting the rest of the domain. Intra-domain correlations of Coh exhibit more a complex pattern to which both secondary (anti-parallel β-strands and β-sheet at the binding interface) and tertiary structure (vicinity of C- and N-termini) contribute. Some of the inter-domain correlations in the complex originate from spatial vicinity and direct interactions, specifically at the Doc:Coh binding interface and at XMod contacts with Doc inserts. However, coupling between distant parts of the complex is also present. For example, fluctuations of the non-binding part of Coh are correlated with the N-terminal part of XMod and strongly anti-correlated with Doc domain.

4 Supplementary Notes

4.1 Constant Barrier Distance Model

The constant barrier distance model[16], also referred to as the Bell-Evans model[22], is commonly used to estimate the distance to the transition state Δx and the natural off-rate k_0 of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments. It predicts that the most probable rupture force $\langle F \rangle$ is linearly dependent on the logarithm of the force loading rate[16]:

$$\langle F(r) \rangle = \frac{k_B T}{\Delta x} \ln \frac{\Delta x \cdot r}{k_0 k_B T} \tag{S1}$$

where k_B is Boltzmann's constant, T is the temperature and r is the loading rate at the point of rupture.

The probability density distribution of rupture forces at given loading rate r in this model is given as[16]:

$$p\,(F) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 \cdot k_B T}{\Delta x \cdot r}\left(e^{\frac{\Delta x}{k_B T} F} - 1\right)\right] \tag{S2}$$

4.2 Dudko-Hummer-Szabo Model

The Dudko-Hummer-Szabo (DHS)[23;24] model describes a non-linear dependence for the most probable rupture force on loading rate:

$$\langle F(r)\rangle = \frac{\Delta G}{\nu \Delta x}\left\{1 - \left[\frac{k_B T}{\Delta G}\ln\left(\frac{k_B T k_0}{\Delta x r}e^{\frac{\Delta G}{k_B T}+\gamma}\right)\right]^{\nu}\right\} \tag{S3}$$

where ΔG is the free energy of activation and $\gamma = 0.577$ is the Euler-Mascheroni constant. The model parameter ν defines the single-well free-energy surface model used ($\nu = \frac{2}{3}$ for linear-cubic and $\frac{1}{2}$ for cusp free-energy. For $\nu = 1$ and $\Delta G \to \infty$ independent of ν the Eqs. (S1) and (S2) are recovered.

4.3 Pearson Correlation and covariance matrix

4.3.1 Validation

An $N \times N$ matrix of Pearson correlation coefficients C_{ij} (Supporting Eq. S4) was calculated from each atom's x, y, z position throughout the simulation trajectory, which inherently ignores off-diagonal elements of the atomic 3×3 submatrices D_{ij}^{mn} from the full normalized $3N \times 3N$ covariance matrix (*i.e.*, correlations along orthogonal axes are neglected, see Supporting Eqs. (S5) and (S6)) and Supporting Fig S8.

Although this quasi-harmonic approximation is commonly employed in correlation analysis[19;25–29], it is not *a priori* justified for complicated biomolecular interactions[30]. To validate the use of Pearson correlations, we therefore first analyzed independently the contributions from diagonal and off-diagonal elements of each 3 x 3 covariance submatrix for each pair of α-carbons within the structure (Fig. S9A and B). Both with and without applied force, the off-diagonal elements roughly follow Gaussian distributions centered around a correlation value of 0. Interestingly, as force was applied, the standard deviation of the distribution of off-diagonal correlation values decreased from $\sigma_{unloaded} = 0.45$ to $\sigma_{loaded} = 0.29$. This indicated a lesser influence of off-diagonal elements on the highly (anti-)correlated motion within the system under force (see Supporting Discussion 3). The diagonal elements of the sub-matrices that are used for calculating the Pearson correlation values showed a dramatically different behavior. Both in the unloaded and loaded state, the resulting distributions were strongly shifted towards highly correlated motion, and the shape of the distribution remained mostly unchanged after application of force. Since our analysis relies on the identification of paths of highest correlation through proximate residues, the quasi-harmonic approximation implied by the use of Pearson correlation is justified, especially for suboptimal pathway analysis. The resulting distributions of on- and off-diagonal matrix elements of each covariance submatrix for the loaded configuration HF class (Fig. S10A) and LF class (Fig. S10B) exhibited the same characteristics as previously described for the native configuration, with off-diagonal elements showing symmetric correlations around zero and diagonal elements showing highly correlated motions.

4.3.2 Supplementary Equations

The Pearson correlation coefficient C_{ij} used in our dynamical network analysis protocol is given by:

$$C_{ij} = \frac{\langle \Delta \mathbf{r}_i(t) \cdot \Delta \mathbf{r}_j(t)\rangle}{\left(\left\langle \Delta \mathbf{r}_i(t)^2\right\rangle \left\langle \Delta \mathbf{r}_j(t)^2\right\rangle\right)^{\frac{1}{2}}} \tag{S4}$$

where $\Delta\mathbf{r}_i(t) = \mathbf{r}_i(t) - \langle\mathbf{r}_i(t)\rangle$.

The full $3N \times 3N$ covariance matrix M_{ij} for atoms i and j consists of 3×3 submatrices of the form:

$$\left\langle \Delta\mathbf{r}_i(t)\,\Delta\mathbf{r}_j(t)^T \right\rangle = M_{ij} = \begin{pmatrix} M_{ij}^{xx} & M_{ij}^{xy} & M_{ij}^{xz} \\ M_{ij}^{yx} & M_{ij}^{yy} & M_{ij}^{yz} \\ M_{ij}^{zx} & M_{ij}^{zy} & M_{ij}^{zz} \end{pmatrix} \tag{S5}$$

The full normalized correlation matrix is calculated from M_{ij}:

$$D_{ij}^{mn} = \frac{M_{ij}^{mn}}{\sqrt{M_{ij}^{mm} M_{ij}^{nn}}} \tag{S6}$$

Consequently, the Pearson correlation coefficient is calculated as the trace of the normalized 3×3 submatrices ($C_{ij} = \mathrm{Tr}\, D_{ij}$).

4.3.3 Derivation of Main Text Equation 2

Eq. 1 from the main text reads:

$$\left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{S7}$$

Combining Eqs. (S7) and (S4) yields:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\left\langle \Delta\mathbf{r}_i^2(t) \right\rangle \left\langle \Delta\mathbf{r}_j^2(t) \right\rangle \right)^{-\frac{1}{2}} \tag{S8}$$

For an arbitrary potential $U_i(\mathbf{r})$ of atom i, a Taylor expansion around the potential minimum (set to be at 0) yields:

$$U_i(\mathbf{r}) = 0 + \underbrace{\mathbf{r}_i^T \nabla U(0)}_{=0} + \frac{1}{2}\mathbf{r}_i^T \underline{H}(0)\,\mathbf{r}_i + ... \tag{S9}$$

where $\underline{H}(0)$ is the Hessian matrix evaluated at the potential minimum. Assuming Schwarz' theorem holds for $U_i(\mathbf{r})$, $\underline{H}(0)$ is a symmetric matrix and therefore has real eigenvalues and orthonormal eigenvectors. Hence, a change to the eigenbasis of $\underline{H}(0)$ is a rotation of the coordinate system. In this new basis the Hessian is diagonal:

$$\underline{H}(0) \to \underline{H}'(0) = \begin{pmatrix} k_{x'} & 0 & 0 \\ 0 & k_{y'} & 0 \\ 0 & 0 & k_{z'} \end{pmatrix} \tag{S10}$$

This yields a simple expression for the second order term in Eq. (S9):

$$U_i(\mathbf{r}') = \frac{1}{2}\mathbf{r}'^T \underline{H}'(0)\,\mathbf{r}' = \frac{1}{2}\left(k_{x'} x'^2 + k_{y'} y'^2 + k_{z'} z'^2 \right) \tag{S11}$$

Now we inspect the normalization of C_{ij}:

$$\langle \Delta\mathbf{r}_i^2(t) \rangle = \langle \mathbf{r}_i^2(t) - 2\mathbf{r}_i(t)\,\langle\mathbf{r}_i(t)\rangle + \langle\mathbf{r}_i(t)\rangle^2 \rangle \tag{S12}$$

In the harmonic approximation of the potential of atom i, $\langle \mathbf{r}_i(t) \rangle = 0$, and therefore $\langle \Delta\mathbf{r}_i(t)^2 \rangle = \langle \mathbf{r}_i^2(t) \rangle$. In the basis of $\underline{H'}(0)$ this becomes:

$$\langle \mathbf{r'}_i^2(t) \rangle = \langle x'_i(t)^2 + y'_i(t)^2 + z'_i(t)^2 \rangle = \langle x'_i(t)^2 \rangle + \langle y'_i(t)^2 \rangle + \langle z'_i(t)^2 \rangle \tag{S13}$$

Applying the equipartition theorem to this result yields:

$$\langle x'_i(t)^2 \rangle = \frac{k_B T}{k'_{x_i}} \tag{S14}$$

And therefore:

$$\langle \Delta\mathbf{r}'_i(t)^2 \rangle = k_B T \left(\frac{1}{k'_{x_i}} + \frac{1}{k'_{y_i}} + \frac{1}{k'_{z_i}} \right) = \frac{k_B T}{k'_{i,eff}} \tag{S15}$$

Plugging this result into Eq. (S8), one finds:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\langle \Delta\mathbf{r}_j(t)^2 \rangle \right)^{-\frac{1}{2}} \tag{S16}$$

Repeating the above steps for atom j yields the final result:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\frac{k_B T}{k'_{j,eff}} \right)^{-\frac{1}{2}} \tag{S17}$$

$$= \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \sqrt{k'_{i,eff} \cdot k'_{j,eff}} \tag{S18}$$

5 Supplementary Figures

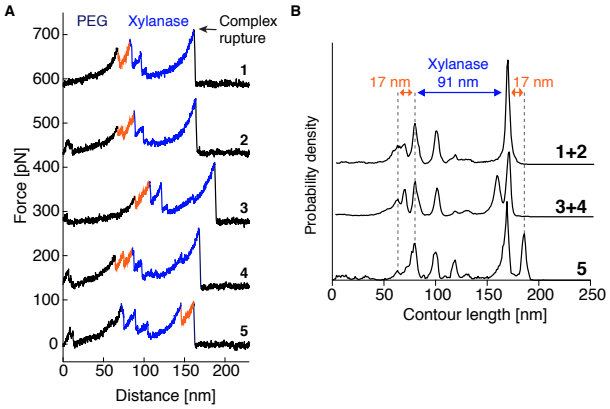


Fig. S1: SMFS of the non-native low force curve class. **A** Typical unfolding fingerprints. All traces showed a characteristic Xyn fingerprint (blue). A $17-19$ nm increment corresponding to partial N-terminal Coh unfolding (orange) occurs either prior to Xyn unfolding (traces 1-4), or just before complex rupture (trace 5). It was observed as a single event (traces 1,3 and 5) or showed substructure (traces 2 and 4). **B** Traces were grouped and assembled into contour length histograms. One or more of the unassigned increments combined into a $17-19$ nm increment.

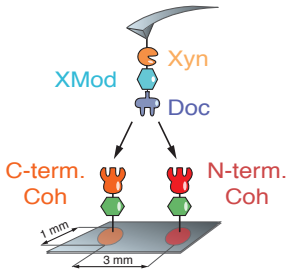


Fig. S2: Comparing the native geometry with the non-native high force class. Two exclude uncertainties in cantilever calibration when comparing the native geometry with the non-native HF class, we immobilized both Coh-CBM (native) and CBM-Coh (non-native) on two spatially separated spots on a single cover glass. These spots where then alternately probed with the same Xyn-XMod-Doc functionalized cantilever.

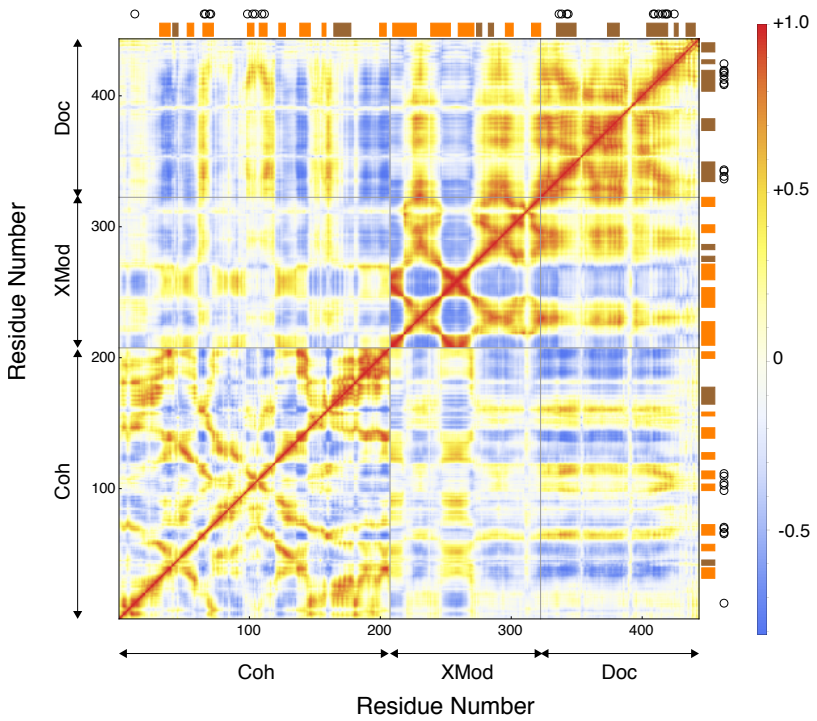


Fig. S3: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the unloaded Xmod-Doc:Coh complex. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from the Coh and Doc binding interface.

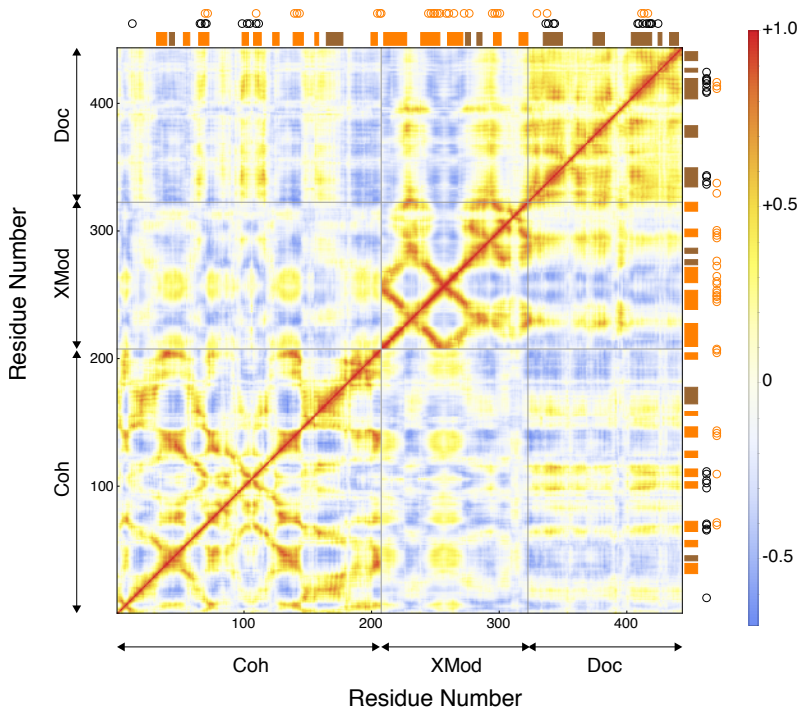


Fig. S4: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the Xmod-Doc:Coh complex loaded with force in the native pulling geometry. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from Coh and Doc binding interfaces and orange circles represent residues on the force propagation path.

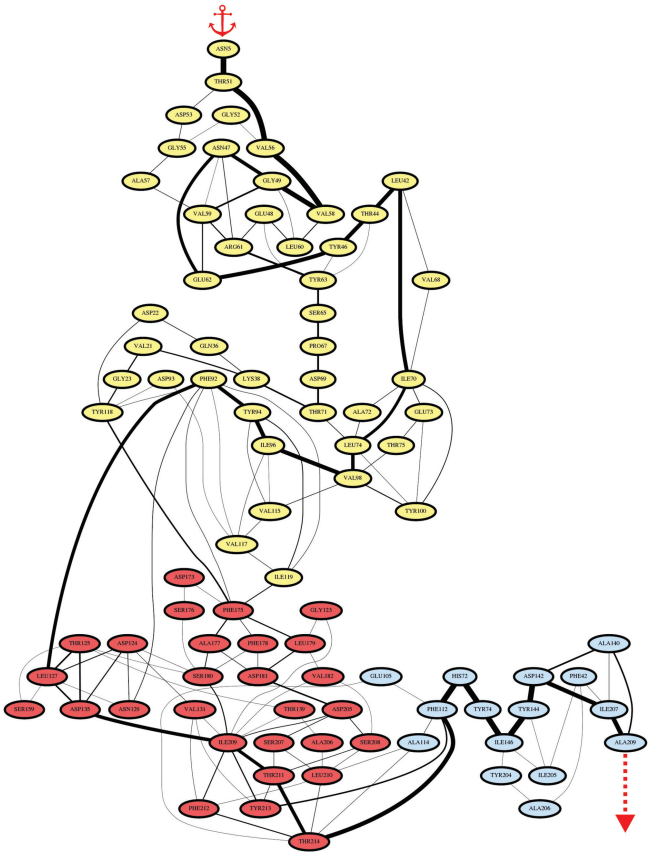


Fig. S5: Force propagation pathway through the loaded XMod-Doc:Coh complex in the native pulling geometry (N-terminal pulling of Xmod-Doc, C-terminal pulling of Coh) obtained from dynamical network analysis. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

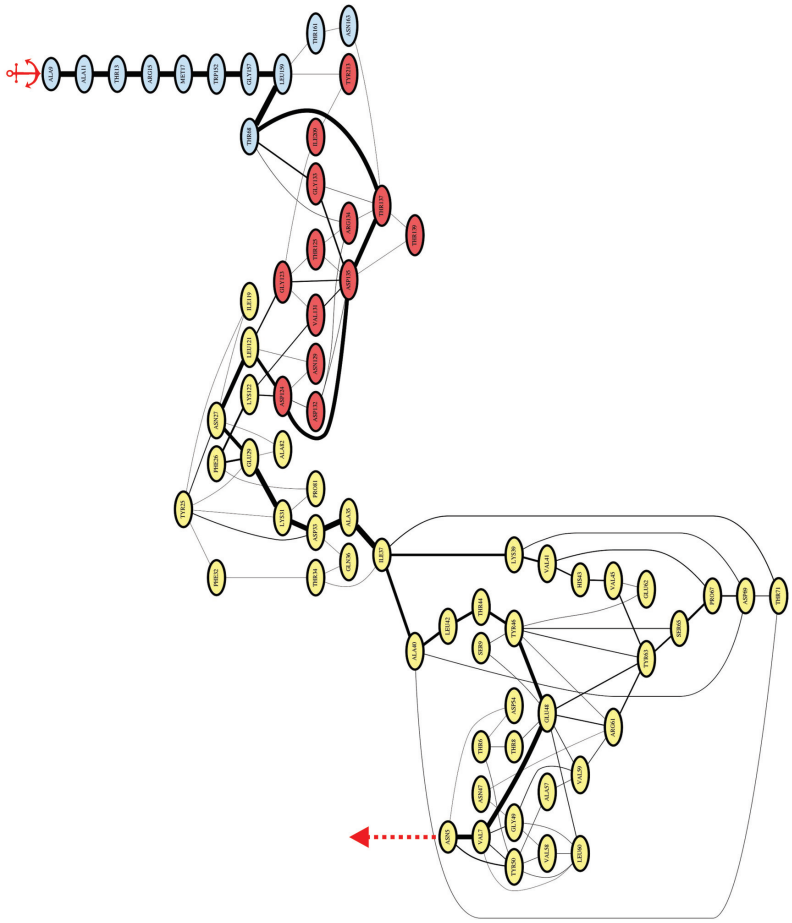


Fig. S6: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing high-force unbinding characteristics and no C-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

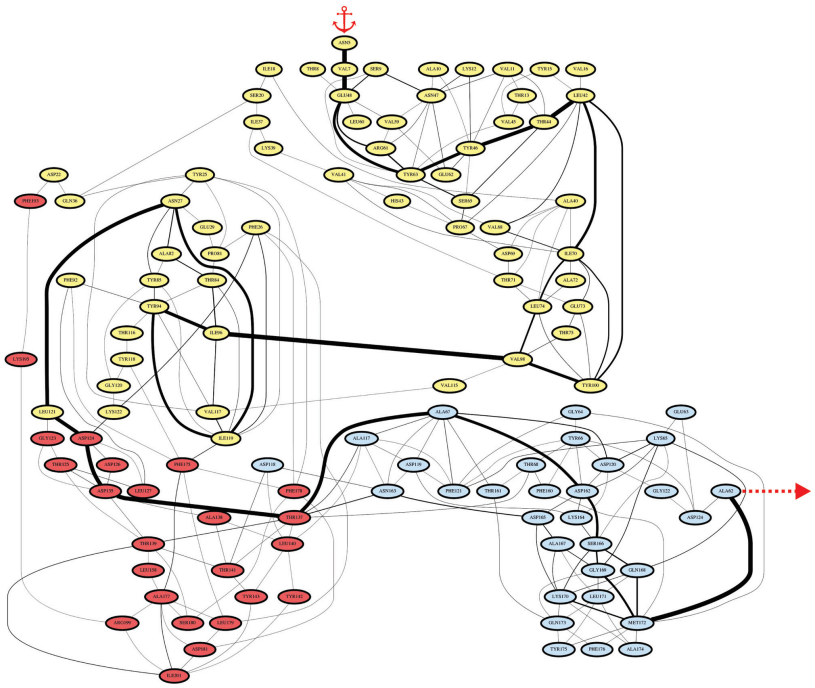


Fig. S7: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing low-force unbinding characteristics and partial N-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

$$M_{ij} =$$

Fig. S8: Full unnormalized covariance Matrix M_{ij} for a five atom system from which the full normalized covariance matrix is calculated according to Eq. (S6). On- and off-diagonal elements from one of the atomic submatrices are highlighted in yellow and blue, respectively.

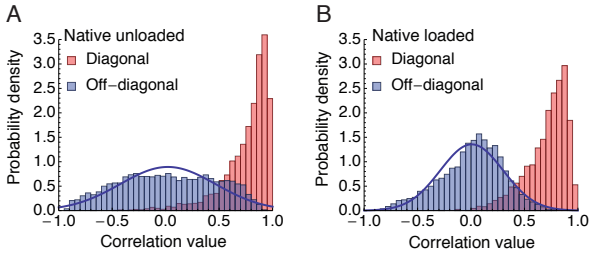


Fig. S9: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A**, the native unloaded, and **B** the native loaded, scenario.

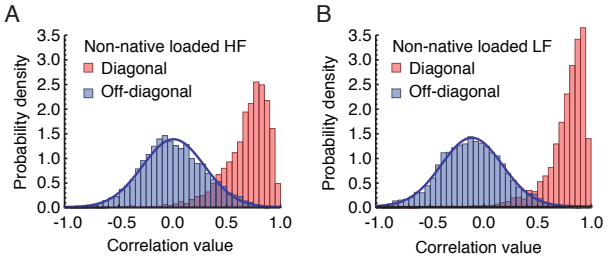


Fig. S10: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A,** the non-native HF, and **B** the non-native LF, scenario.

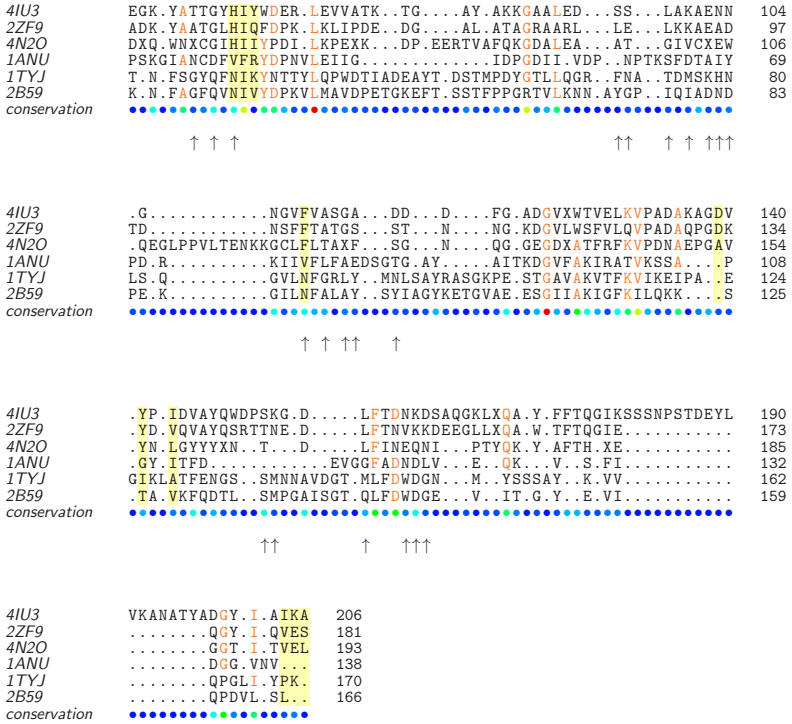


Fig. S11: Structure-aligned sequences of six crystallized cohesins. Residues on the force propagation path are highlighted in yellow. Arrows indicate binding residues. Residue conservation is color-coded from blue - lack of conservation, to red - residue fully conserved. Crystal structures used: 4IU3 ScaE *Rf* FD-1, 2ZF9 ScaE *Rf* strain 17, 4N2O CohG *Rf* FD-1, 1ANU CohC2 CipC *Ct*, 1TYJ CohA11 ScaA *Bc*, 2B59 SdbA *Ct*.

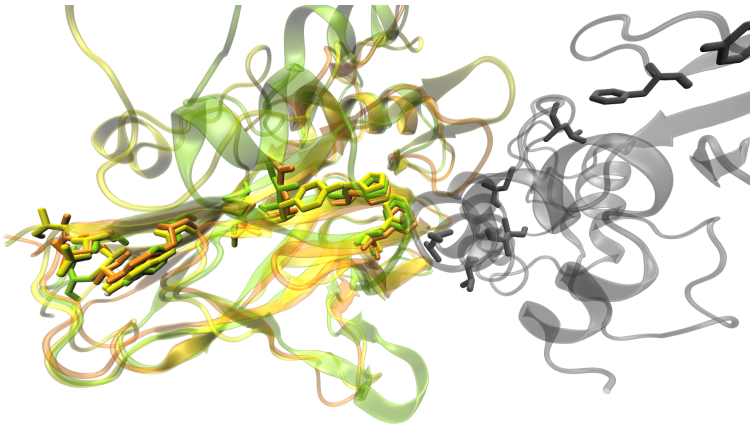


Fig. S12: Structure and sequence conservation of the force propagation pathway residues in Coh. CohE from the ScaE cell anchoring protein, *Rf* FD-1 used in this work (PDB 4IU3) is highlighted in green. Highly homologous structures of CohE from *Rf* strain 17 (PDB 2ZF9) and Coh G from *Rf* FD-1 (PDB 4A2O) are colored in orange and yellow, respectively. Residues lying in the force propagation path are shown as sticks. XDoc from the CttA *Rf* FD-1 scaffold used in this work is shown in gray.

References

[1] Orly Salama-Alber, Maroor K Jobby, Seth Chitayat, Steven P Smith, Bryan A White, Linda J W Shimon, Raphael Lamed, Felix Frolow, and Edward A Bayer. Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.*, 288(23):16827–16838, April 2013.

[2] Constantin Schoeler, Klara H Malinowska, Rafael C Bernardi, Lukas F Milles, Markus A Jobst, Ellis Durner, Wolfgang Ott, Daniel B Fried, Edward A Bayer, Klaus Schulten, Hermann E Gaub, and Michael A Nash. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.*, 5:1–8, December 2014.

[3] F William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.*, 41(1):207–234, May 2005.

[4] Markus A Jobst, Constantin Schoeler, and Michael A Nash. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.*, 82(82):e50950, 2013.

[5] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.*, 1(1):280–285, June 2006.

[6] Stefan W Stahl, Michael A Nash, Daniel B Fried, Michal Slutzki, Yoav Barak, Edward A Bayer, and Hermann E Gaub. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.*, 109(50):20431–20436, December 2012.

[7] Elias M Puchner, Gereon Franzen, Mathias Gautel, and Hermann E Gaub. Comparing proteins by their unfolding pattern. *Biophys. J.*, 95(1):426–434, July 2008.

[8] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[9] R C Bernardi, M C R Melo, and K Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta*, 1850(5):872–877, 2015.

[10] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.*, 151(1): 283–312, 1999.

[11] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[12] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro E M Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell, Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1and χ 2Dihedral Angles. *J. Chem. Theory Comput.*, 8(9):3257–3273, 2012.

[13] A D MacKerell, D Bashford, Bellott, R L Dunbrack, J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiórkiewicz-Kuczera, D Yin, and M Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[14] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926, 1983.

[15] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[16] S Izrailev, S Stepaniants, M Balsera, Y Oono, and K Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72(4):1568–1581, 1997.

[17] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[18] João V Ribeiro, Juan A C Tamames, Nuno M F S A Cerqueira, Pedro A Fernandes, and Maria J Ramos. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug. Des.*, 82(6):743–755, December 2013.

[19] J Eargle and Z A Luthey-Schulten. NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics*, 28(22):3000–3001, 2012.

[20] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 106(16):6620–6625, April 2009.

[21] N Glykos. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.*, 27(14):1765–1768, 2006.

[22] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85(1):5–15, July 2003.

[23] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, March 2006.

[24] O K Dudko, G Hummer, and A Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *PNAS*, 105(41):15755–15760, October 2008.

[25] M Karplus and J N Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.

[26] R M Levy, M Karplus, J Kushick, and D Perahia. Evaluation of the Configurational entropy for Proteins: Application to Molecular Dynamics Simulations of an α-Helix. *Macromolecules*, 17(7):1370–1374, 1984.

[27] Rommie E Amaro, Anurag Sethi, Rebecca S Myers, V Jo Davisson, and Zaida A Luthey-Schulten. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry*, 46(8):2156–2173, February 2007.

[28] Rebecca W Alexander, John Eargle, and Zaida Luthey-Schulten. Experimental and computational determination of tRNA dynamics. *FEBS Lett.*, 584(2):376–386, January 2010.

[29] Rafael C Bernardi, Isaac Cann, and Klaus Schulten. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels*, 7(83):1–8, 2014.

[30] Oliver F Lange and Helmut J Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, March 2006.

4.4 Associated Publication P4

Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy

by

Markus A. Jobst, Lukas F. Milles, Constantin Schoeler, Wolfgang Ott, Daniel B. Fried, Edward A. Bayer, Hermann E. Gaub and Michael A. Nash

published in

eLIFE, 4, e10319, (2015)

eLIFE elifesciences.org

RESEARCH ARTICLE

Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy

Markus A Jobst[1,2], Lukas F Milles[1,2], Constantin Schoeler[1,2], Wolfgang Ott[1,2], Daniel B Fried[3], Edward A Bayer[4], Hermann E Gaub[1,2], Michael A Nash[1,2]*

[1]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-University, Munich, Germany; [2]Center for Nanoscience, Ludwig-Maximilians-University, Munich, Germany; [3]Kean University, New Jersey, United States; [4]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel

Abstract Receptor-ligand pairs are ordinarily thought to interact through a lock and key mechanism, where a unique molecular conformation is formed upon binding. Contrary to this paradigm, cellulosomal cohesin-dockerin (Coh-Doc) pairs are believed to interact through redundant dual binding modes consisting of two distinct conformations. Here, we combined site-directed mutagenesis and single-molecule force spectroscopy (SMFS) to study the unbinding of Coh:Doc complexes under force. We designed Doc mutations to knock out each binding mode, and compared their single-molecule unfolding patterns as they were dissociated from Coh using an atomic force microscope (AFM) cantilever. Although average bulk measurements were unable to resolve the differences in Doc binding modes due to the similarity of the interactions, with a single-molecule method we were able to discriminate the two modes based on distinct differences in their mechanical properties. We conclude that under native conditions wild-type Doc from *Clostridium thermocellum* exocellulase Cel48S populates both binding modes with similar probabilities. Given the vast number of Doc domains with predicteddual binding modes across multiple bacterial species, our approach opens up newpossibilities for understanding assembly and catalytic properties of a broadrange of multi-enzyme complexes.
DOI: 10.7554/eLife.10319.001

***For correspondence:** michael.nash@lmu.de

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 23 July 2015
Accepted: 28 October 2015
Published: 31 October 2015

Reviewing editor: Taekjip Ha, Johns Hopkins University School of Medicine, United States

Introduction

Cellulosomes are hierarchically branching protein networks developed by nature for efficient deconstruction of lignocellulosic biomass. These enzyme complexes incorporate catalytic domains, carbohydrate binding modules (CBMs), cohesin:dockerin (Coh:Doc) pairs, and other conserved features (***Demain et al., 2005***; ***Bayer et al., 2004***; ***Schwarz, 2001***; ***Béguin and Aubert, 1994***; ***Smith and Bayer, 2013***; ***Fontes and Gilbert, 2010***). A central attribute of cellulosome assembly is the conserved ~75 amino acid type-I Doc domain typically found at the C-terminus of cellulosomal catalytic domains. The highly conserved consensus Doc sequence from *Clostridium thermocellum* (*Ct*) is shown in ***Figure 1A***. Dockerins guide attachment of enzymes into the networks by binding strongly to conserved Coh domains organized within non-catalytic poly (Coh) scaffolds. In addition to their nanomolar binding affinities, many archetypal Coh:Doc pairs are thought to exhibit dual binding modes (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***; ***Currie et al., 2012***). The bound Doc domain can adopt two possible orientations that differ by ~180˚ rotation on the Coh surface, as shown in ***Figure 1B***. The two binding modes originate from duplicated F-hand sequence motifs, a conserved structural feature found among type-I dockerins (***Pagès et al., 1997***). The duplicated F-hand motifs resemble EF-hands found in eukaryotic calcium binding proteins (e.g., calmodulin), and provide

eLife digest Some bacteria use cellulose, the main component of plant cell walls, as a food source. The enzymes that break down cellulose are anchored onto a protein scaffold in a structure called the cellulosome on the bacteria's surface. This anchoring occurs through an interaction between receptor proteins known as 'cohesin' domains on the scaffold proteins and 'dockerin' ligands on the enzymes.

Most receptor-ligand interactions only allow the two proteins to bind in a single, fixed orientation. However, cohesins and dockerins are suspected to bind in two different configurations. It has been difficult to investigate the populations of these different configurations because most experimental techniques investigating protein binding take average measurements from many molecules at once. As the binding modes are extremely similar, these methods have been unable to distinguish between the two cohesin-dockerin binding configurations without introducing mutations, in part because these configurations are very similar to each other.

Jobst et al. used a technique called single-molecule force spectroscopy to investigate cohesin-dockerin interactions between individual molecules. This technique applies a force that separates, or 'unbinds', cohesin and dockerin, by pulling individual complexes of the two binding partners apart with a nanoscale probe. In the experiments, *E. coli* bacteria were made to produce mutant versions of dockerin that can only bind to cohesin in one orientation. This allowed each binding configuration to be studied individually. The results of these experiments revealed the mechanical unbinding patterns of each cohesin-dockerin configuration, and showed that it is possible to use these patterns to distinguish between the two configurations. A complimentary set of experiments revealed that wild-type (non-mutated) cohesin-dockerin complexes occupy both configurations in approximately equal amounts, and do not switch modes once bound.

Further single-molecule experiments together with computer simulations will provide a more detailed picture of how cohesin and dockerin fit together in the two configurations. Such experiments could also reveal how cohesin and dockerin contribute to the break down of cellulose inside living cells and how they could be used for the precise assembly of single proteins.

DOI: 10.7554/eLife.10319.002

internal sequence and structural symmetry to Doc domains. Rotating Doc by ~180˚ with respect to Coh (***Figure 1B,C***) results in an alternatively bound complex with similarly high affinity involving the same residues on Coh recognizing mirrored residues within Doc. The dual binding mode is thought to increase the conformational space available to densely packed enzymes on protein scaffolds, and to facilitate substrate recognition by catalytic domains within cellulosomal networks (***Bayer et al., 2004***). From an evolutionary perspective, the dual binding mode confers robustness against loss-of-function mutations, while allowing mutations within Doc to explore inter-bacterial species cohesin-binding promiscuity in cellulosome-producing microbial communities. Coh:Doc interactions and dual binding modes are therefore important in the context of cellulose degradation by cellulosome-producing anaerobic bacterial communities.

However, direct experimental observation of the dual binding modes for wild-type Doc has thus far proven challenging. Ensemble average bulk biochemical assays (e.g., surface plasmon resonance, calorimetry, enzyme-linked immunosorbent assays) are of limited use in resolving binding mode populations, particularly when the binding modes are of equal thermodynamic affinity. Crystallography is challenging because the complex does not adopt a unique molecular conformation, but rather exhibits a mixture of two conformations thereby hindering crystal growth. Structural data on the dual binding mode have typically been collected using a mutagenesis approach, where one of the binding modes was destabilized by mutating key recognition elements (***Carvalho et al., 2007***; ***Pinheiro et al., 2008***). This approach, however, while resolving the structures of each bound complex, cannot determine if one binding mode is dominant for wild-type Doc, or if that dominance is species or sequence dependent. Coarse grained molecular dynamics has also predicted dual modes of interaction between Coh and Doc (***Hall and Sansom, 2009***), but direct experimental evidence of both binding modes for wild-type Doc has remained elusive. Improved fundamental understanding of the dual binding mode could shed light onto the molecular mechanisms by which these multi-

eLIFE Research article Biochemistry | Biophysics and structural biology

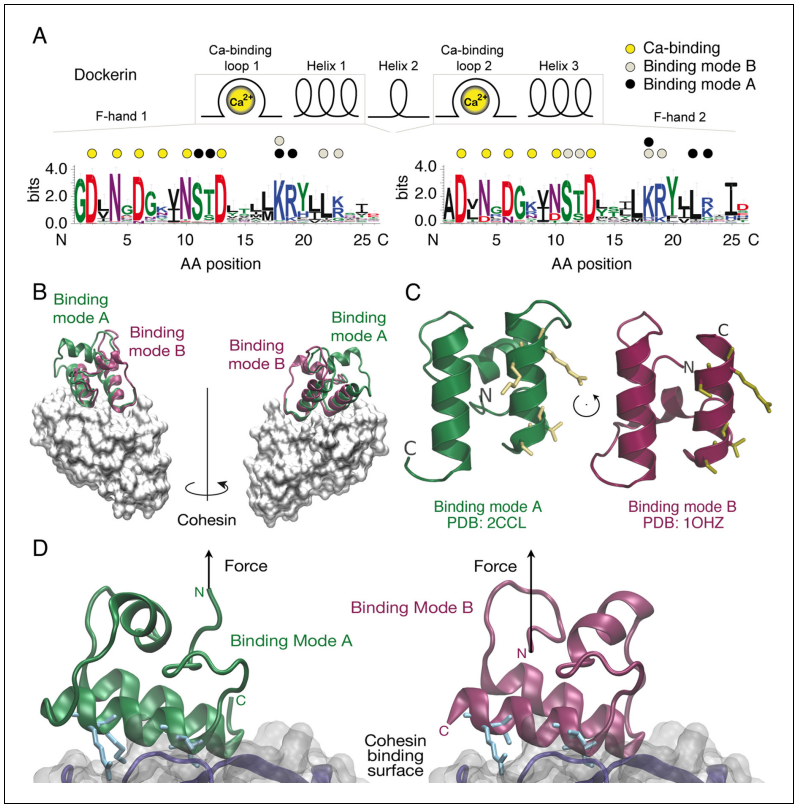


Figure 1. Cohesin:Dockerin dual binding modes. (**A**) Secondary structure and consensus sequence logo (***Crooks, 2004***) assembled from 65 putative *Ct* type-I Doc variants. Dots above the amino acid codes indicate residues involved in: Ca^{2+} coordination (yellow), mode A binding (black), and mode B binding (gray). Letter colors represent chemical properties: Green, polar; purple, neutral; blue, basic; red, acidic; black, hydrophobic. Crucial Coh-binding residues are located at positions 11, 12, 18, 19, 22, and 23 in each F-hand motif. (**B**) Coh:Doc complex crystal structures showing overlaid Doc domains in the two binding modes. Images were generated by aligning the Coh domain (gray) from PDB 2CCL (green, binding mode (**A**) and 1OHZ (red, binding mode (**B**) using the VMD plugin MultiSeq (***Humphrey et al., 1996***; ***Roberts et al., 2006***). (**C**) View of the Doc binding interface for each mode from the perspective of Coh. The conserved binding residues at positions 11, 12, 18, and 19 in the F-hand motif relevant for binding in the corresponding mode are depicted as stick models (yellow). (**D**) Close-up view of the interface for each binding mode with arrows indicating the location and direction of applied force. Binding residues 11, 12, 18, and 19 for binding mode A and 45, 46, 52, and 53 for binding mode B are shown as blue stick models. The Coh domain is oriented the exact same way in both views.

DOI: 10.7554/eLife.10319.003

enzyme complexes self-assemble and achieve synergistic conformations, as well as provide a new approach to designing systems for protein nanoassembly (***Kufer et al., 2009***; ***2008***).

Here, we used SMFS (***Li and Cao, 2010***; ***Engel and Müller, 2000***; ***Woodside and Block, 2014***) to study wild-type and mutant Doc from exocellulase Cel48S of *C. thermocellum* (*Ct*-DocS). We demonstrate that specific unfolding/unbinding trajectories of individually bound Coh:Doc complexes

are characteristic of the binding modes. To validate our approach, we produced Doc mutants that exhibited a preferred binding mode. We performed single-molecule pulling experiments on bound Coh:mutant Doc complexes and observed a strong bias in the probability of two clearly distinguishable unfolding patterns, termed 'single' and 'double' rupture types for each binding mode mutant. We further probed the unbinding mechanism of the double rupture events using poly (Gly-Ser) inserts to add amino acid sequence length to specific sections of Doc as a means to identify which portions of Doc unfolded. Finally, we used the inherent differences in mechanical stability of each binding mode, and the effects these differences had on the unfolding force distributions of an adjacent domain, to directly observe and quantify binding mode populations for wild-type Doc.

Results

Protein design

The wild-type and mutant Doc sequences used in this work were aligned (***Beitz, 2000***) and are presented in ***Figure 2***. Among *Ct*-Doc domains, a Ser-Thr pair located at positions 11 and 12 of F-hand motif 1 (N-terminal helix 1) is highly conserved (***Figure 1A***). This Ser-Thr pair is H-bonded to Coh in binding mode A (***Figure 1A***, black dots). Analogously, binding mode B refers to the configuration where the Ser-Thr pair from helix 3 dominates the H-bonding to Coh (***Figure 1A***, gray dots). Binding mode B was previously crystallized for a homologous *Ct*-Doc (***Carvalho et al., 2003***). Mutation of the Ser-Thr pair in helix 3 to Ala-Ala was used to bias binding and thereby crystallize binding mode A for the same Doc (***Carvalho et al., 2007***). A similar targeted mutagenesis approach was also used to obtain crystal structures of a *Clostridium cellulolyticum* Doc in each binding mode (***Pinheiro et al., 2008***).

To preferentially select for a specific binding mode (A or B), we prepared Doc sequences that incorporated 4 amino acid point mutations, referred to as quadruple mutants ('Q'). To design quadruple mutants, we noted that recent structural work reported a set of *Ct*-Doc domains that differ from the canonical duplicated Ser-Thr sequences. These non-canonical Docs were found to exhibit only a single binding mode (***Brás et al., 2012***; ***Pinheiro et al., 2009***). In one of these non-canonical Doc domains, an Asp-Glu pair was found in place of Ser-Thr. Since the Coh surface is negatively charged, we postulated that including Asp-Glu in place of Ser-Thr within one of the F-hands could be used to effectively knock out a given binding mode for our canonical Doc. Additionally, we incorporated double alanine mutations to replace the conserved Lys-18 Arg-19 residues of a given F-hand motif, further destabilizing a targeted binding mode. Q1 refers to a quadruple mutant where helix 1 has been modified at four positions (i.e. S11D-T12E-K18A-R19A). Q3 refers to the quadruple mutant where helix 3 has been modified at four positions (i.e. S43D-T44E-K50A-R51A). As a negative

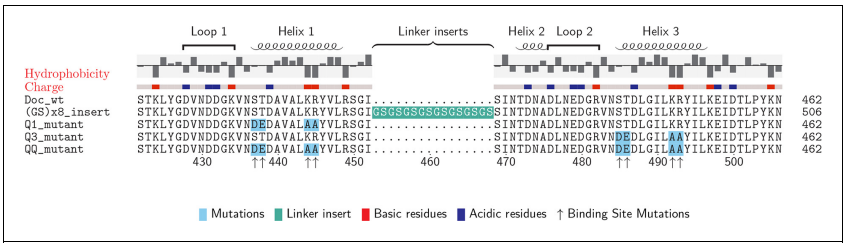


Figure 2. Doc sequences used in this study (N- to C-terminus). Doc_wt: wild-type sequence; hydrophobicity and charge graphs are displayed for the wild-type-Doc (red: positively charged, blue: negatively charged); (GS)x8_insert: A $(Gly\text{-}Ser)_8$ linker was incorporated between helix 1 and helix 2; Q1_mutant: Quadruple mutant in helix 1. Four point mutations (DE/AA) were incorporated into Doc helix 1 to knock out binding mode A; Q3_mutant: Quadruple mutant in helix 3. Four point mutations (DE/AA) were incorporated into Doc helix 3 to knock out binding mode B; QQ_mutant: Non-binding control with both binding modes knocked out. Numbers below indicate amino acid number of the fusion protein construct starting from the xylanase N-terminus.

DOI: 10.7554/eLife.10319.004

control, we prepared a mutant referred to as 'QQ' that incorporated quadruple mutations into both helices 1 and 3.

Doc domains were expressed as fusion domains attached to the C-terminal end of xylanaseT6 (Xyn) from *Geobacillus stearothermophilus* to improve solubility and expression levels as previously reported (***Stahl et al., 2012***). The Xyn domain also acts as a so-called fingerprint in AFM force extension traces to provide a means for screening datasets and searching for known contour length increments. We use the term 'contour length' to refer to the maximum length of a stretched (unfolded) polypeptide chain. Our screening process identified single-molecule interactions and ensured correct pulling geometry. For the Coh domain, we chose cohesin 2 from *Ct*-CipA expressed as a C-terminal fusion domain with the family 3a carbohydrate binding module (CBM) from *Ct*-CipA. In order to exclude artifacts arising from fingerprint domains, protein immobilization or pulling geometry, a second set of fusion proteins was cloned, expressed and probed in complementary experiments using a flavoprotein domain from the plant blue light receptor phototropin (iLOV) (***Chapman et al., 2008***). All protein sequences are provided in the 'Materials and methods' section.

Single-molecule unfolding patterns

The pulling configuration for single-molecule AFM experiments is shown in ***Figure 3A***. CBM-Coh was site-specifically and covalently attached to an AFM cantilever tip and brought into contact with a glass surface modified with Xyn-Doc. The mechanical strength of protein domains and complexes will strongly depend on the pulling points (i.e. sites at which the molecule is attached to cantilever/surface). The site-specific attachment chemistry used here was precisely defined by the chosen residue of immobilization, ensuring the same loading geometry was used on the complex for each and every data trace. After formation of the Coh:Doc complex, the cantilever was retracted at a constant speed that ranged from 200 to 3200 nm/s while the force was monitored by optical cantilever deflection. The resulting force-distance traces were characteristic of the series of energy barriers crossed by the protein complex along the unfolding/unbinding pathway. A sawtooth pattern was consistently observed when molecular ligand-receptor complexes had formed. Sorting the data using contour length transformation (***Puchner et al., 2008***) and identifying traces that contained a Xyn contour length increment (~89 nm) allowed us to screen for single-molecule interactions (***Stahl et al., 2012***), as described in our prior work on Coh:Doc dissociation under force (***Stahl et al., 2012***; ***Schoeler et al., 2014***; ***Jobst et al., 2013***; ***Otten et al., 2014***; ***Schoeler et al., 2015***).

Typical single-molecule interaction traces from such an experiment are shown in ***Figure 3B, C*** and in ***Figure 3—figure supplement 1***. Following PEG linker stretching, an initial set of peaks

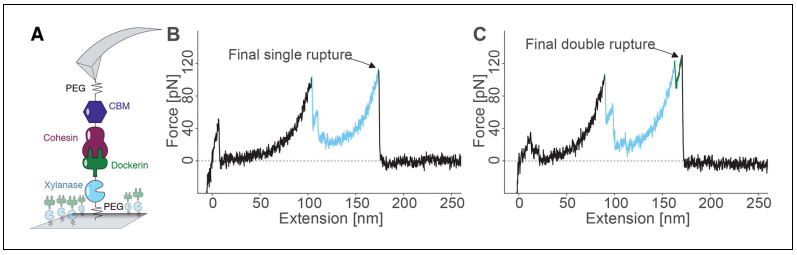


Figure 3. Overview of the experimental configuration and recorded single-molecule unfolding and unbinding traces. (**A**) Schematic depiction showing the pulling geometry with CBM-Coh on the AFM Cantilever and Xyn-Doc on the glass substrate. Each fusion protein is site-specifically and covalently immobilized on a PEG-coated surface. (**B-C**) Each force vs. extension trace shows PEG linker stretching (black), xylanase unfolding and subsequent stretching (blue), and Coh:Doc complex rupture. The Coh:Doc complex rupture occurred in two distinct event types: single (**B**) and double (**C**) ruptures. The 8-nm contour length increment separating the double peaks was assigned to Doc unfolding (C, green).

DOI: 10.7554/eLife.10319.005

The following figure supplement is available for figure 3:

Figure supplement 1. Representative sample of force traces.

DOI: 10.7554/eLife.10319.006

sequentially decreasing in force was assigned to xylanase unfolding and stretching. This domain when unfolded added ~89 nm of free contour length to the system. The final peak (s) corresponded to rupture of the Coh:Doc complex, and occurred as either 'single' or 'double' rupture events. The contour length increment between the two double event peaks was found to be ~8 nm, that is, 8 nm of hidden contour length was added to the biopolymer during a sub-step of Doc unbinding (see 'Discussion'). The 8-nm contour length increment was also observed in complementary experiments employing other fusion domains: xylanase was swapped for an sfGFP domain and CBM was swapped out for an iLOV domain. In these new fusions, the 8 nm Doc increment was still observed, indicating it was not caused by a specific fusion domain. As we show below, double and single rupture events were associated with binding modes A and B, respectively. CBM unfolding length increments (~57 nm) were only rarely observed because the Coh:Doc complex only rarely withstood forces sufficiently high to unfold CBM (***Stahl et al., 2012***).

Ensemble average binding experiments

Binding experiments were carried out in bulk to evaluate the binding affinity of wild-type, Q1, Q3, and QQ Doc sequences to wild-type Coh. Xyn-Doc fusion protein variants were immobilized in a microwell plate and exposed to tag red fluorescent protein (TagRFP) (***Merzlyak et al., 2007***) fused to Coh (TagRFP-Coh) across a range of concentrations, followed by rinsing and subsequent fluorescence readout (***Figure 4A***). The data clearly showed that Q1 and Q3 Doc sequences, each with a mutated binding mode, maintained high-binding affinity with dissociation constants (K_d) in the nM range. These values are in good agreement with previous reports on homologous type-I Doc domains (***Brás et al., 2012***; ***Sakka et al., 2011***). This suggested that mutant Doc domains with one destabilized binding mode were still able to recognize fluorescent protein fused Coh with strong affinity by relying on the alternative binding mode that was preserved. The QQ double knockout mutant, however, showed no appreciable binding over the concentration range tested. This negative control showed that DEAA quadruple mutations were in fact effective at eliminating binding for the targeted modes.

Single-molecule rupture statistics of binding mode mutants

For each Doc tested, we collected tens of thousands of force-extension traces and selected for further analysis only those traces showing the ~89 nm xylanase contour length increments and no other anomalous behavior, resulting in typically 200–3000 usable single-molecule interaction curves per experiment. We determined the number of Coh:Doc unbinding events that occurred as single or double rupture peaks. The results are shown in ***Figure 4B***. The wild-type Doc showed double rupture events in ~57% of the cases, and single rupture events in ~43% of the cases. The mutant designed to knock out binding mode A (Q1), showed a single event probability of ~77%, and a double event probability of ~23%. The mutant designed to knock out binding mode B (Q3) showed a single event probability of ~41%, and a double event probability of ~59%. It is clear from these data that the Q1 mutant has a strong bias toward single peaks that is not observed in the wild-type leading to preliminary assignment of single peaks to binding mode B.

For all double events, we determined the force difference of the second peak relative to the first (***Figure 4C***). Q1 and wild-type on average showed second peaks that were ~15–20% higher in force than the first peak. Q3 meanwhile showed clearly different behavior. Although the ratios of single to double peaks were nearly identical between wild-type and Q3, differences in the relative force between the first and second peaks differentiated wild-type and Q3 (***Figure 4C***). Double peaks for the Q3 mutant were more likely to show a shielded behavior, where the second peak was lower in force than the first peak by ~10%. Although the Q3 mutant showed the same single vs. double event probability as wild-type, the double events for Q3 were distinguishable from those of the wild-type based on this observed decrease in the rupture force of the second peak. The second barrier of the double events was therefore weaker in Q3 than for wild-type. This weaker 2^{nd} double peak for the Q3 mutant combined with similar single/double peak ratios as wild-type leads us to believe that the number of double peaks is being underestimated systematically for the Q3 mutant. Generally, each binding mode still allows for the occurrence of a single event (albeit with different likelihood), in which the whole Doc domain unbinds without an additional unfolding substep. Since the second and final energy barrier for complex dissociation is weaker than the first for the Q3 mutant, the

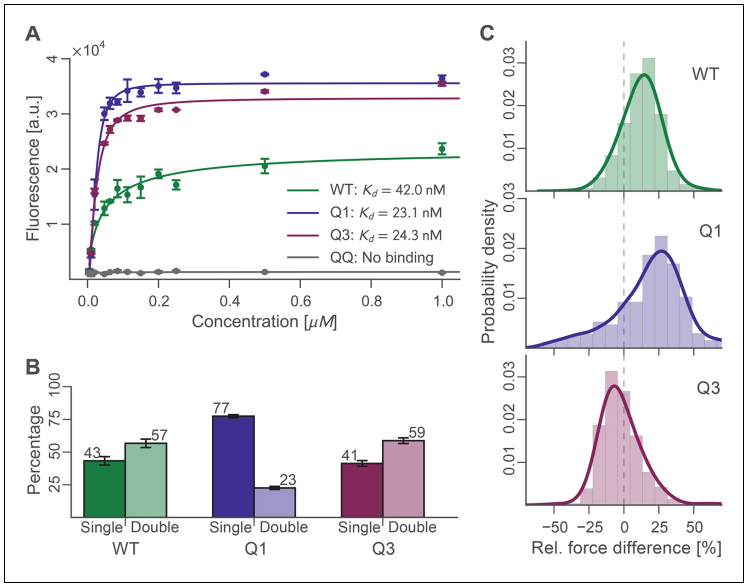


Figure 4. Bulk and single-molecule characterization of Doc mutants. (**A**) Fluorescence binding curve showing binding of TagRFP-labelled Coh to wild-type and mutant Doc nonspecifically immobilized in a 96-well plate. Both Q1 and Q3 mutants bound TagRFP-Coh similarly to wild-type with dissociation constants (K_D) in the low nM range. The negative control QQ mutant showed no binding. Solid lines are 4 parameter logistic nonlinear regression model fits to the data. Error bars represent the standard deviation of three independent samples. (**B**) Event probabilities for single (opaque colors) and double (translucent colors) Coh:Doc rupture peaks determined for Doc wild-type and DE/AA quadruple mutants. Data originate from 947, 4959, and 1998 force-extension traces from wild-type, Q1 and Q3 variants, respectively. Error bars represent 95% Clopper-Pearson confidence intervals based on the beta probability distribution. (**C**) Relative difference in double peak rupture forces for the different Doc variants. Positive values indicate a stronger final peak. Histograms represent concatenated data from various pulling speeds. Drawn lines are kernel density estimates calculated on the raw data.

DOI: 10.7554/eLife.10319.007

The following source data is available for figure 4:

Source data 1. Probability Data.

DOI: 10.7554/eLife.10319.008

probability for the molecule to pass both barriers simultaneously is increased, thus resulting in a higher percentage of single events.

Probing the 8-nm length increment with poly (GS) inserts

We sought to identify the molecular origin of the 8 nm contour length increment separating the double event peaks by engineering additional amino acid sequence length into the Doc domain. Amino acid insert sequences have previously been used to probe length increments in AFM force spectroscopy experiments (***Bertz and Rief, 2009***) (***Carrion-Vazquez et al., 1999***). By adding additional amino acids to the polypeptide chain at a particular location, insert sequences increase the gain in contour length following unfolding of a subdomain in a predictable way. Any change in the observed length increment can be pinpointed to the position in the molecule where the unfolding event occurs. In this case, we engineered flexible $(GS)_8$ insert sequences directly into wild-type Doc between helices 1 and 2, in a flexible loop that was not expected to interfere with either of the two binding modes. Structural homology models (***Figure 5A***) of the wild-type Doc and $(GS)_8$ insert

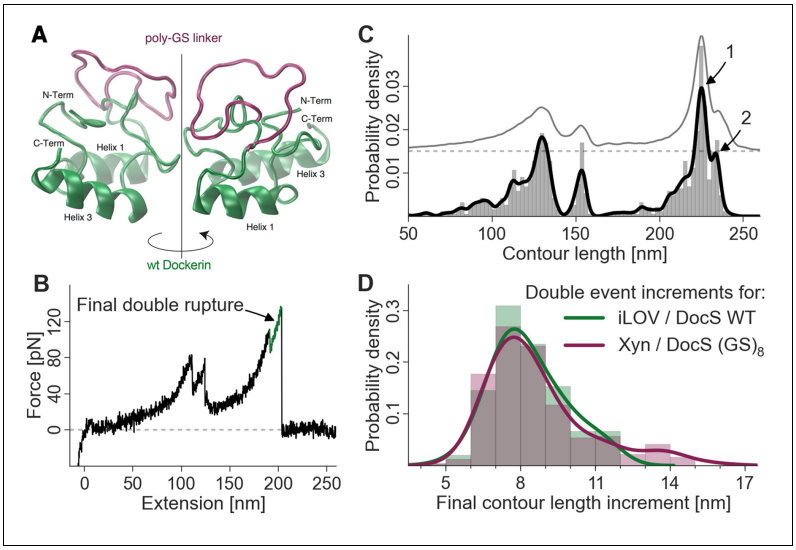


Figure 5. Probing the final contour length increment with Poly (GS) inserts. (**A**) Structural homology model overlay of wild-type and mutant Doc containing a $(GS)_8$-linker between helix 1 and helix 3. The wild-type Doc is shown in green. The 16 amino acid long GS-insert is shown in purple (***Kelley and Sternberg, 2009***) (remaining Doc domain not shown). (**B**) Typical force extension trace with final double rupture event depicted in green (arrow). (**C**) Histogram and kernel density estimate of the transformation of the single force extension trace in panel B into contour length space (black) and kernel density estimate of the whole dataset of single molecule Xyn-Doc:Coh-CBM traces bearing xylanase fingerprint and final double rupture (gray, offset in y-direction for readability) in contour length space. (**D**) Histograms (bars, bin width: 1 nm), kernel density estimates (drawn lines, bandwidth: 0.75 nm, gaussian kernel), and statistical test (Kolmogorov-Smirnov, 'KS test') are each calculated on the raw data of the final increments (peak-to-peak distances) in contour length space (x-distance between arrow 1 and 2 in panel (**C**). Maxima for final double event increments lie at 7.75 nm and 7.73 nm for iLOV-Coh:Doc (wild-type)-sfGFP (N = 255) and Xyn-Doc $(GS)_8$:Coh-CBM (N = 320) final ruptures, respectively (a two-sample KS test on the raw data indicates no significant difference in the data distributions (p-value of 21.7%).

DOI: 10.7554/eLife.10319.009

sequence were calculated using the Phyre server (***Kelley and Sternberg, 2009***). If the 8-nm contour length increment was caused by sequential unbinding of Doc helices 1 and 3 in wild-type Doc, then double peaks for the poly (GS) constructs should show an increase in the double peak contour length increment. As shown in ***Figure 5B,C and D***, the contour length histogram for $(GS)_8$ Doc was indistinguishable from the wild-type Doc. No additional contour length was gained due to additional amino acids inserted between Doc helices 1 and 2. Since the Doc was anchored to the glass slide through an N-terminal xylanase domain, this result indicated that the unfolding event responsible for the 8-nm length increment must be located upstream (i.e. N-terminal) from the site of the $(GS)_8$-insert. This result suggested that unfolding of calcium binding loop 1 and helix 1 in Doc was the source of the 8-nm length increment.

Single-molecule evidence of dual binding mode

To finally confirm the presence of both bound conformations in wild-type Coh:Doc complexes, we replaced xylanase with sfGFPand CBM with iLOV as the contour length marker or fingerprint domains. iLOV was chosen as a superior unfolding fingerprint domain because it does not show multiple unfolding substeps (in contrast to xylanase), which simplified analysis. Also iLOV has an unfolding force distribution that lies in a similar range as the Coh:Doc complex dissociation single and

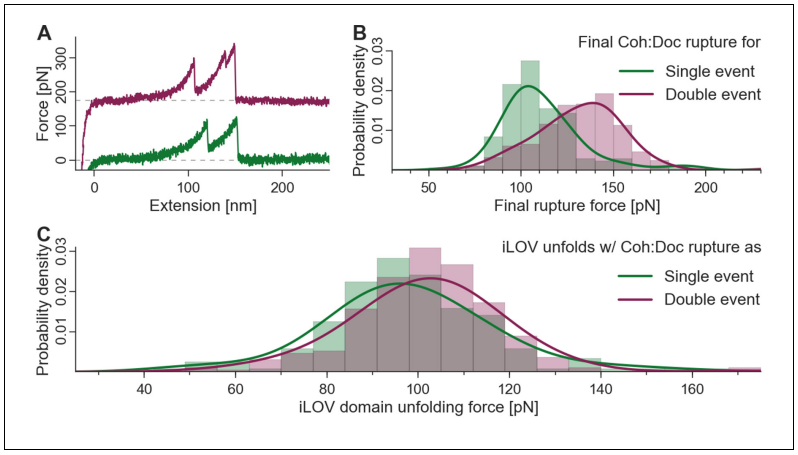


Figure 6. Biasing of unfolding force distributions by dual binding mode. (**A**) Typical force traces showing iLOV unfolding with final single (green) and double (purple) complex ruptures. The curve terminating in a double peak is offset in the y-direction for clarity. (**B**) Final complex rupture force distribution for single and double events. Double events are more mechanically stable. (**C**) iLOV domain unfolding forces for final single (green) and double (red) events at a pulling velocity of 800 nm/s. Histograms (bars), kernel density estimates (lines), and statistical tests are each obtained from the raw data. Maxima for iLOV unfolding lie at 96.0 pN and 102.7 pN for single (N = 172) and double (N = 277) final ruptures, respectively. A two-sample Kolmogorov-Smirnov test showed significant differences in the data distributions (p-value of 0.09%). Since the data were all recorded with a single cantilever and both event types were distributed equally throughout the runtime of the measurement, no systematic biasing is expected. Because of the lower force distribution of final single peaks, the iLOV unfolding force distribution is truncated compared to final double peak force traces, supporting the notion that the binding mode is set prior to mechanical loading of the complex.

DOI: 10.7554/eLife.10319.010

double peaks, allowing for effective biasing of the iLOV unfolding force distributions by the inherent stability difference between single and double event peaks. ***Figure 6A*** shows characteristic single and double event curves containing iLOV unfolding (36-nm contour length increment) followed by Coh:Doc rupture as a single or double event. The rupture force distributions of the single and double event (second peak) ruptures are shown in ***Figure 6B***. The most probable rupture force for single events was ~104 pN, while for double events this value was ~140 pN at a pulling speed of 800 nm/s. We next calculated the unfolding force distributions of the iLOV domain for curves that terminated with single events or double events. If the Coh:Doc complex ruptured before iLOV unfolding was observed, the curve was eliminated from the dataset because it lacked a fingerprint domain length increment. This criterion for inclusion in the dataset results in a biasing of the iLOV unfolding forces, since the maximum of the fingerprint unfolding force distribution that can be observed must lie below that of the Coh:Doc complex. The fact that we observed a downward shift in the iLOV unfolding forces (***Figure 6C***) for curves that terminated in the less mechanically stable single rupture event is confirmation that the single- and double-event peaks arise from separate bound conformations. Each mode has a distinct mechanical stability and energy landscape that is set at the time of receptor-ligand binding, that is once bound, the conformation of the complex does not change. If single- and double-event unbinding patterns were simply two competing pathways out of the same bound state, then the downward shift in rupture force distribution would not be observed for the iLOV unfolding forces. Although this shift in rupture force distributions is comparatively subtle, it can be observed accurately with high statistical significance. We note that the datasets for both binding modes were measured with the same cantilever throughout the runtime of the whole experiment. Calibration and drift issues therefore did not interfere with the required accuracy.

Discussion

The relatively small ~8 kDa Doc domains exhibit an internal sequence and structural symmetry that is believed to give rise to a dual mode of binding to Coh, as shown in ***Figure 1***. In order to study this remarkable plasticity in molecular recognition in greater detail, we prepared a series of mutants (***Figure 2***) designed to either knock out a specific binding mode or add length to the molecule at a specific position. Bulk experiments showed that Doc mutants Q1 and Q3, originally designed to suppress one of the binding modes, were still able to bind Coh with high affinity, while the double knockout did not bind (***Figure 4A***). The equilibrium affinities of Coh binding to Q1, Q3, or wild-type were all similarly high with K_Ds in the low nM range, in good agreement with literature values (***Sakka et al., 2011***), suggesting the two binding modes are thermodynamically equivalent and rendering them indistinguishable with conventional methods such as ELISA or calorimetry. Techniques like surface plasmon resonance could possibly show differing values for on- and off-rates for the mutants, but would still not be able to resolve the binding modes within a wild-type population.

Force spectroscopy with the AFM interrogates individual molecules, and measures their mechanical response to applied force. Since the technique is able to probe individual members of an ensemble, it provided a means to quantify binding mode configurations by assigning unfolding/unbinding patterns to the binding mode adopted by the individual complexes. Site-directed Q1 and Q3 mutations supported the assignment of binding mode A to a characteristic double rupture peak dissociation pathway. Single events were assigned to binding mode B and showed no Doc unfolding substep prior to complex rupture.

We consistently observed 8 nm of added contour length that separated the Doc double peaks. Since force is applied to Doc from the N-terminus, we analyzed the Doc sequence starting at the N-terminus and searched for reasonable portions of Doc that could unfold in a coordinated fashion to provide 8 nm of contour length. The results from the GS-insert experiments (***Figure 5***) indicated no change in the double-event contour length increment, regardless of the added GS-insert length located between helix 1 and 3 in Doc. This result is consistent with the 8 nm length increment being located N-terminally from the GS-insert site, implicating unfolding of Doc calcium binding loop 1 and helix 1 as the source of the 8 nm. This length accurately matches the estimated length increment for unfolding calculated from the crystal structure (***Figure 1D***).

Although this result could also be consistent with the 8 nm increment being located somewhere outside the Doc domain in the polyprotein, we deem this scenario highly unlikely. The 8 nm increment cannot be located in the Xyn or CBM domains because we have accounted for Xyn and CBM lengths in their entirety based on the observed 89 nm and rare 57 nm length increments here and in a previous study (***Stahl et al., 2012***), and for confirmation swapped out those domains for different proteins completely (i.e. iLOV and GFP). The remaining possibility that the 8 nm is located within the Coh domain is also not likely since the barrel-like structure of the Coh is known to be mechanically highly stable (***Valbuena et al., 2009***; ***Hoffmann et al., 2013***). Also, if the 8-nm length increment were due to partial Coh unfolding, the Q1 and Q3 mutants would not be expected to affect the single/double peak ratio or force differences between the double event peaks as was observed (***Figure 4B, C***). The GS-insert data suggest the 8-nm length increment is located within Doc, upstream (N-terminal) from the GS-insert site implicating calcium loop 1 and helix 1 in this unfolding event.

Finally, we observed that an inherent difference in the mechanical stability of single and double event rupture peaks (***Figure 6B***) could be used as a feature by which to discriminate the binding modes. Our analysis algorithm accepted only the force curves that first showed iLOV fingerprint domain unfolding followed by either a single- or double-rupture peak. By observing a small but significant downward shift in the iLOV unfolding force distribution when analyzing curves that terminated in the less stable single-event peak, we confirmed the single-event peaks originate from a unique conformation that is 'set' at the time of complex formation.

Taken together, we propose an unbinding mechanism where the first barrier of the double peaks represents unfolding of the N-terminal calcium binding loop and unraveling of alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19 in the wild-type structure in binding mode A. Based on a length per stretched amino acid of 0.4 nm, the expected contour length for unfolding the Doc domain up to this position would be 7.6 nm, in good agreement with the measured value of 8 nm within experimental error. A portion of the N-terminal calcium binding loop (i.e. residues S11-

T12) is involved in binding to D39 in Coh. The first peak of the double events is attributed to breakage of this interaction and simultaneous unfolding of calcium loop 1 and alpha helix 1 up to the Lys-Arg pair at sequence positions 18 and 19. Another contributing factor is the intramolecular clasp that has been identified as a stabilizing mechanism among similar type-I Doc domains (**Slutzki et al., 2013**). A recent NMR structural study (**Chen et al., 2014**) on the same wild-type Doc used in this work confirmed a hydrophobic ring-stacking interaction between Tyr-5 and Pro-66. Confirmation of this clasp motif by NMR means the head and tail of the Doc are bound together, additionally stabilizing the barrier that is overcome in the first of the double event peaks. In this scenario, subsequent to breaking the interactions between the calcium binding loop and Coh, disrupting the intramolecular clasp and unfolding the N-terminal loop-helix motif, the remaining bound residues including Lys-18, Arg-19, Lys-50, Leu-54, and Lys-55 stay bound to Coh and are able to withstand substantial force on their own, eventually breaking in the second and final of the double rupture peaks. Prior work further supports this unbinding mechanism, revealing that a progressive N-terminal truncation of Doc did not affect the interaction largely, unless the truncation reached the Lys-18 and Arg-19 residues (**Karpol et al., 2009**). This corroborates the idea of the C-terminal end of helix 1 being a crucial part of the binding site within the complex. Single rupture peaks were thus observed when the wild-type complex was bound in binding mode B, and no unfolding of Ca-binding loop 1 or helix 1 occurred. Force was propagated directly to bound residues Lys-18, Leu-22, and Arg-23 which when broken resulted in complete complex dissociation.

Given the fingerprint biasing phenomenon (**Figure 6C**), we finally sought to correct the single/double peak counting statistics (**Figure 4B**) in order to correct for undercounting of single peaks due solely to their failure to reach sufficiently high forces to unfold the fingerprint domain. Only traces showing a fingerprint were analyzed to ensure defined unfolding geometry. Using the rupture force distributions of singles, doubles, iLOV, and xylanase domains, we calculated the probability of occurrence of fingerprint unfolding at a force higher than the single-event ruptures. This overlap probability was found to be 0.85 for iLOV and 0.40 for xylanase. When the single/double peak ratios for were corrected for this effect, the final binding mode ratios (binding mode A/binding mode B, i.e., doubles/singles) were found to be 0.95 and 0.87 for xylanase-Doc and iLOV, respectively. These ratios are close to 1 indicating comparable probability of each binding mode after accounting for biasing the single/double peak counting statistics due to fingerprint domain stability. We note that these numbers are also slightly lower than unity due to the exclusion of double peaks that occurred before unfolding of the fingerprint domains. Further details on rupture force distributions and overlap statistics are shown in **Figure 7**. As the magnitude of biasing changes with the unfolding force distributions of each fingerprint domain, overlaps in the probability distributions allow for normalizing single/double event ratios of experimental data sets with different fingerprinting domains. For the Coh:Doc complex unbinding event, biasing (undercounting) is more pronounced for the mechanically weaker single ruptures. This normalization procedure shows the relative difference of biasing between single and double events, as double events are less biased than single events.

The biological significance of Coh-Doc interactions in the context of cellulosome assembly and catalysis cannot be overstated. Their high affinity and specificity, along with their modularity, thermostability, and their ultrastable mechanical properties all make Coh-Doc unique from a biophysics perspective, and attractive from an engineering standpoint. Dual binding mode Doc domains are broadly predicted among many cellulosome producing bacteria (e.g. *C. thermocellum, C. cellulolyticum, R. flavefaciens*), however relatively few have been confirmed experimentally (**Carvalho et al., 2007**; **Pinheiro et al., 2008**; **Brás et al., 2012**). In fact, the direct effect of single vs. dual binding modes on the ability of cellulosomes to convert substrate into sugars is currently unknown. It is therefore unclear whether or not dual binding modes affect, for example, the catalytic properties of native or engineered synthetic cellulosomes.

However, it is important to note that cellulosome producing bacteria invariably live among communities with other microorganisms, which may be producing cellulases and cellulosomes of their own. In such an environment, a dual binding mode could enable organisms to produce enzymes that are able to bind to a neighboring species' scaffoldins, yet still retain high-affinity interactions with host scaffoldins. They would be able to combine resources with neighboring cells in a mixed microbial consortium. The dual binding mode could therefore allow genetic drift to explore interspecies protein binding. Indeed, cross-species reactivity between Coh and Doc has been reported (**Haimovitz et al., 2008**). Cellulosome-producing microbes may therefore be pursuing a middle

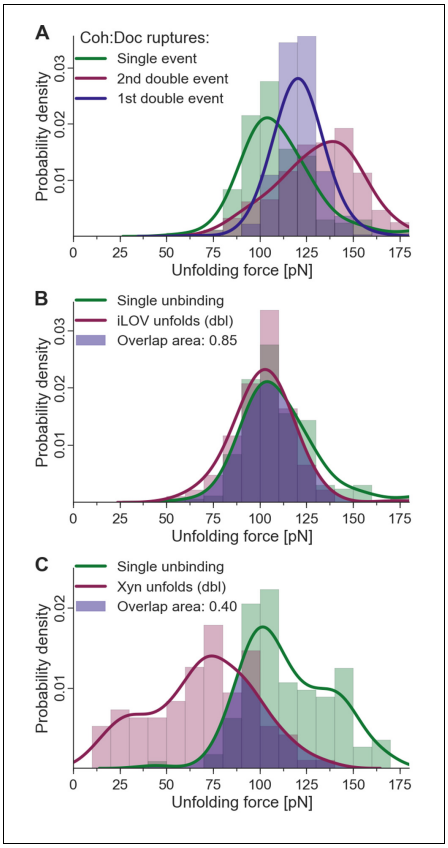


Figure 7. Fingerprint unfolding and complex unbinding forces. (**A**) Rupture force distribution of final complex ruptures for single (green), first (purple) and second (red) double unbinding events. (**B**) Overlap area (purple) of iLOV domain unfolding force distribution (red) (iLOV-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. (**C**) Overlap area (purple) of Xyn domain unfolding force distribution (red) (Xyn-doubles curve class) with the rupture force distribution (green) for single-event complex ruptures. Overlaps in probability distributions allow normalizing single-event counts to double events to account for different biasing caused by the different unfolding forces of the fingerprint domain. Biasing occurs, because for overlapping force distributions of fingerprint unfolding and complex ruptures, unbinding events are more likely to take place without fingerprint unfolding if the two distributions are closer together. For the Coh:Doc unbinding, this effect is more pronounced for the weaker single ruptures. Because double events are also biased, this still does not give a true quantification, but only compensates for the differences of biasing. The non-bell-evans-like shape of the single rupture peaks, especially in the region of the 1st double event peak (**A**) suggests that this class of curves does not contain a single type of unbinding mechanism, but rather a superposition of different event types.

DOI: 10.7554/eLife.10319.011

ground between protein synthesis strictly for selfish vs. communal usage. By distinguishing the presence of each binding mode for wild-type Doc domains, the single-molecule biophysical approach presented here based on differences in mechanical hierarchies will facilitate further study into the significance of the dual binding mode.

In summary, the dual binding mode of Coh:Doc domains has so far proven resistant to explicit experimental characterization. Crystallography combined with mutagenesis has provided snapshots of the two modes, but resolving each of the modes for wild-type Doc under near native conditions has up until now not been possible. We have demonstrated the advantages of a single-molecule approach in resolving these subtle differences in molecular conformations of bound complexes. Despite having equal thermodynamic binding affinity, when mechanically dissociated by pulling from the N-terminus of Doc, binding mode A was more mechanically stable with an additional energy barrier. This mechanical difference was exploited to probe the two binding modes independently from one another, providing direct observation of this unique mechanism in molecular recognition. In the future, harnessing control over binding modes could offer new approaches to designing molecular assembly systems that achieve defined protein orientations.

Materials and methods

Site-directed mutagenesis of plasmid DNA

A pET28a vector containing the previously cloned xylanaseT6 from *Geobacillus stearothermophilus* (***Salama-Alber et al., 2013***) and DocS dockerin from *Clostridium thermocellum* Cel48S were subjected to QuikChange mutagenesis (***Wang and Malcolm, 1999***) to install the following mutations: Q1, Q3, and QQ in the dockerin and T129C in the xylanase, respectively.

For insertion of the $(GS)_4$ and $(GS)_8$ linkers into the Doc domain, exponential amplification with primers bearing coding sequences for the inserts at their 5′-ends was performed with a Phusion High-Fidelity DNA polymerase (New England Biolabs, MA). PCR products were then blunt end ligated using KLD Enzyme Mix and KLD Reaction Buffer from the Q5 site directed mutagenesis kit (New England Biolabs, MA). The modified DNA constructs were used to transform *Escherichia coli* DH5-alpha cells, grown on kanamycin-containing agar plates and subsequently screened. All mutagenesis products were confirmed by DNA sequencing analysis.

Primers used for inserting the $(GS)_8$ linker into the Doc domain:

Fw 5′-ggttctggctccggttctggctccagcatcaacactgacaat-3′

Rev 5′-agaaccggagccagagccggaacctatacctgatctcaaaacatatct-3′

Protein expression and purification

Fusion proteins HIS-CBM A2C-Coh2 (*C.t.*) were expressed in *E. coli* BL21(DE3)RIPL cells in kanamycin-containing media supplemented with 2mM calcium chloride overnight at 16˚C. After harvesting, cells were lysed by sonication, and the lysate was subjected to heat treatment at 60˚C for 30 min to precipitate the bulk of the host bacterial proteins, leaving the expressed thermophilic proteins in solution. The lysate was then pelleted, and the supernatant fluids were applied to a beaded cellulose column and incubated at 4˚C for 1 hr. The column was then washed with 50 mM Tris buffer (pH 7.4) containing 1.15 M NaCl, and the protein was eluted using a 1% (vl/v) triethylamine aqueous solution. Tris buffer was added to the eluent and the solution was neutralized with HCl.

Fusion proteins HIS-Xyn T129C-DocS (*C.t.*) wild-type, Q1, and Q3 mutants were expressed as described above. Following heat treatment, the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20mM imidazole and 2mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was then dialyzed to remove the imidazole.

Fusion proteins ybbR-HIS-CBM A2C-Coh2 (*C.t.*), ybbR-HIS-Xyn T129C-DocS (*C.t.*) wild-type and QQ mutants and ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_4$ insert were expressed in *E. coli* BL21(DE3) RIPL cells; ybbR-HIS-Xyn T129C-DocS (*C.t.*) $(GS)_8$ insert fusion protein variants were expressed in *E. coli* NiCo21(DE3)RIPL cells. Cultivation and expression was done in ZYM-5052 autoinduction media (***Studier, 2005***) containing kanamycin (and chloramphenicol, in case of the NiCo21(DE3)RIPL cells) overnight at 22˚C, overall 24 hr. After harvesting, cells were lysed using sonication. The lysate was then pelleted by centrifugation at 39,000 rcf, the supernatant fluids were applied to Ni-NTA columns

and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with polyacrylamide gravity flow columns or with polyacrylamide spin desalting columns.

All protein solutions were concentrated with Amicon centrifugal filter devices and stored in 50% (v/v) glycerol at -20˚C (ybbR-free constructs) or -80˚C (ybbR-bearing constructs). The concentrations of the protein stock solutions were determined to be in the order of 1–15 mg/mL by absorption spectrophotometry at a wavelength of 280 nm.

ELISA-like binding assay

1 µM of Xyn-Doc fusion proteins (wild-type Q1, Q3, QQ Doc fusions) bearing either wild-type or mutant Doc domains were adsorbed onto surfaces of the wells of a 96-well nunc maxi sorp plate (Thermo Scientific, Pittsburgh, PA). After blocking (2% (w/v) BSA, 0.05% Tween 20 in TBS buffer) and several rinsing steps, a red fluorescent protein-cohesin (StrepII-TagRFP-Coh2 (*C.t.*), Addgene ID 58,710 (***Otten et al., 2014***)) fusion construct was incubated to the unspecifically immobilized Doc fusion proteins over a range of concentrations. After further rinsing, the fluorescence of the TagRFP domain was measured with a multi-well fluorescence plate reader (M1000 PRO, Tecan Group Ltd., Männedorf, Switzerland). Fluorescence values were plotted against their corresponding concentration values for each protein variant, and 4 parameter logistic nonlinear regression model functions were fitted to the data to determine the transition point of the curve.

Surface immobilization strategies

The Xyn domain had a cysteine point mutation at position 129 (Xyn T129C) to facilitate covalent attachment to a glass surface *via* Polyethylene glycol (PEG)-maleimide linkers. There were no other cysteines within the Xyn or Doc domains, which ensured site-specific immobilization of the molecule and defined mechanical loading of Doc from the N-terminus for the AFM experiments. The CBM domain likewise contained an A2C cysteine point mutation for covalent attachment to the cantilever tip *via* PEG-maleimide linkers. The second set of fusion proteins sfGFP-Doc and iLOV-Coh was covalently attached to coenzyme A bearing PEG linkers by their terminal ybbR tags.

AFM sample preparation

For AFM measurements, silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation nominal spring constant: 100 pN/nm; 25 kHz resonance frequency in water), and glass coverslips (Menzel Gläser, Braunschweig, Germany; diameter 22mm) were used. 3-Aminopropyl dimethyl ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany), α-Maleinimidohexanoic-ω-NHS PEG (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany; PEG-MW: 5 kDa), immobilized tris (2-carboxyethyl)phosphine (TCEP) disulfide reducing gel (Thermo Scientific, Pittsburgh, PA), tris (hydroxymethyl) aminomethane (TRIS, >99% p.a., Carl Roth, Karlsruhe, Germany), $CaCl_2$ (>99% p.a., Carl Roth, Karlsruhe, Germany), sodium borate (>99.8% p.a., Carl Roth, Karlsruhe, Germany), NaCl (>99.5% p.a., Carl Roth, Karlsruhe, Germany), ethanol (>99% p.a.), toluene (>99.5% p.a., Carl Roth, Karlsruhe, Germany) were used as received. Sodium borate buffer was 150 mM, pH 8.5. Measurement buffer for AFM-SMFS was tris-buffered saline supplemented with 1 mM $CaCl_2$ (TBS, 25 mM TRIS, 75 mM NaCl, 1 mM $CaCl_2$ pH 7.2). All buffers were filtered through a sterile 0.2 µm polyethersulfone membrane filter (Nalgene, Rochester, NY) prior to use.

Force spectroscopy measurement samples, measurements and data analysis were prepared and performed according to previously published protocols (***Jobst et al., 2013***;***Otten et al., 2014***). In brief, NHS-PEG-Maleimide linkers were covalently attached to cleaned and amino-silanized silicon nitride AFM cantilevers and cover glasses. The respective protein constructs were covalently linked either *via* engineered cysteine residues to the maleimide groups of the surface on the sample directly, or *via* Sfp phosphopantetheinyl transferase-mediated attachment of a terminal ybbR tag to coenzyme A, which was previously attached to the maleimide groups of the surface.

AFM-SMFS measurements

AFM data were recorded in 25 mM TRIS pH 7.2, 75 mM NaCl and 1mM $CaCl_2$ buffer solution (TBS). Retraction velocities for constant speed force spectroscopy measurements varied between 0.2 and 3.2 µm/s. Cantilever spring constants were calibrated utilizing the thermal method applying the

equipartition theorem to the one dimensionally oscillating lever (**Hutter and Bechhoefer, 1993**; **Cook et al., 2006**). Measurements were performed on custom built instruments, deploying an Asylum Research (Santa Barbara, CA, USA) MFP-3D AFM controller and Physik Instrumente (Karlsruhe, Germany) or attocube (Munich, Germany) piezo nanopositioners (**Gumpp et al., 2009**). After each measurement, the xy-stage was actuated by 100 nm to probe a new spot on the surface and measure new individual Xyn-Doc fusion molecules. Instrument control software was programmed in Igor Pro 6.3 (Wavemetrics). The retraction speed was controlled with a closed-loop feedback system running internally on the AFM controller field-programmable gate array (FPGA).

Force-extension data analysis

Data analysis and plotting was performed in Python (Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org) utilizing the libraries NumPy and SciPy (**van der Walt et al., 2011**) and Matplotlib (**Hunter, 2007**).

Measured raw data were analyzed by determining the zero force value with the baseline position and applying a cantilever bending correction to the z-position. The resulting force distance traces were coarsely screened for peaks as sudden drops in force and curves with less than three peaks (such as in ***Figure 3—figure supplement 1***, panel F) were excluded, as they contain no clearly identifiable signal. Force-distance traces were transformed into contour length space with the inverse worm-like-chain model (**Jobst et al., 2013**), assuming a fixed persistence length of 0.4 nm. Screening for the 89 nm xylanase, the 36nm iLOV and the final 8 nm final double rupture increment was performed by finding their corresponding local maxima in a kernel density estimate with bandwidth b = 1 nm. Thresholds in force, distance, and peak counts were applied to sort out nonspecific and multiple interactions. All curves were ultimately selected for the xylanase or iLOV fingerprint and checked manually. For the counting statistics, double peaks were detected as an increment of 8 +- 4 nm in contour length for final rupture peaks in the contour length plot, given that the curve showed one of the fingerprints. If a double peak was detected, the force difference was determined as the percentual difference between the first and the final rupture peak force.

Barrier position diagrams were assembled using optimal alignment through cross-correlation (**Puchner et al., 2008**; **Otten et al., 2014**). The numbers of points included in fitted histograms are provided in the figure captions, along with the statistical tests and significance values obtained.

Amino acid sequences

pET28a-HIS-XynT129C-DocS (*C.t.*) wild-type

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALK-
RYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-XynT129C-DocS (*C.t.*) 16aa GS Insert

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAV-
EPYQLQNEKDVQMLKRHFNSIVAENVMKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVP-
QWFFLDKEGKPMVNECDPVKREQNKQLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWY-
QIAGIDYIKVAFQAARKYGGDNIKLYMNDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSE-
AEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFW-
GIADNHTWLDSRADVYYDANGNVVVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGT-
PSTKLYGDVNDDGKVNSTDAVALKRYVLRSGIGSGSGSGSGSGSGSGSSINTDNADLNEDGRVNSTDLGI-
LKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q1 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-

QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNEDEAVALA-
AYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) Q3 mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNSTDAVALK-
RYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-HIS-XynT129C-DocS (*C.t.*) QQ mutant

MSHHHHHHKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQMLKRHFNSIVAENV-
MKPISIQPEEGKFNFEQADRIVKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNK-
QLLLKRLETHIKTIVERYKDDIKYWDVVNEVVGDDGKLRNSPWYQIAGIDYIKVAFQAARKYGGDNIKLYM-
NDYNTEVEPKRTALYNLVKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSM-
YGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYEKLSDKISNVTFWGIADNHTWLDSRADVYYDANGNV-
VVDPNAPYAKVEKGKGKDAPFVFGPDYKVKPAYWAIIDHKVVPGTPSTKLYGDVNDDGKVNEDEAVALA-
AYVLRSGISINTDNADLNEDGRVNDEDLGILAAYILKEIDTLPYKN

pET28a-ybbR-HIS-sfGFP-DocIS (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHHPWTSASSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEG-
DATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYK-
TRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQL-
ADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKKVVPGTPST-
KLYGDVNDDGKVNSTDAVALKRYVLRSGISINTDNADLNEDGRVNSTDLGILKRYILKEIDTLPYKN

pET28a-ybbR-HIS-CBM A2C-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHHPWTSASMCNTVSGNLKVEFYNSNPSDTTNSINPQFKVTNT-
GSSAIDLSKLTLRYYYTVDGQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTG-
GTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPSTQP-
VTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEII-
GIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFAD-
NDLVEQKVSFIDGGVNVGNAT

pET28a-ybbR-HIS-iLOV-Coh2 (*C.t.*)

MGTDSLEFIASKLALEVLFQGPLQHHHHHHHPWTSASGSPEFIEKNFVITDPRLPDNPIIFASDGFLELTEYSR-
EEILGRNARFLQGPETDQATVQKIRDAIRDQRETTVQLINYTKSGKKFWNLLHLQPVRDQKGELQYFIGV-
QLDGSDHVGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSK-
GIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATV-
KSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

pET28a-StrepII-TagRFP-Coh2 (*C.t.*)

MWSHPQFEKVSKGEELIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDI-
LATSFMYGSRTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFPS-
NGPVMQKKTLGWEANTEMLYPADGGLEGRSDMALKLVGGGHLICNFKTTYRSKKPAKNLKMPGVYYVD-
HRLERIKEADKETYVEQHEVAVARYCDLPSKLGHKLNGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAGSD-
GVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTAIYPDRKII-
VFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVNVGNAT

Acknowledgements

The authors acknowledge Carlos Fontes, Sarah Teichmann, Stefan Stahl, and Ellis Durner for helpful discussions. Support for this work was provided by the ERC Advanced Grant CelluFuel, and the EU 7th Framework Programme NMP4- SL-2013-604530 (CellulosomePlus), and the German-Israeli Foundation (GIF) for Scientific Research and Development. MAN acknowledges support from Society in Science – The Branco Weiss Fellowship from ETH Zurich.

Additional information

Funding

Funder	Grant reference number	Author
European Research Council	294438	Hermann E Gaub
European Commission	NMP4- SL-2013-604530	Daniel B Fried
German-Israeli Foundation for Scientific Research and Development	G-147-207.4-2012	Edward A Bayer Hermann E Gaub Michael A Nash
Society in Science	Branco Weiss Fellowship	Michael A Nash

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MAJ, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article, Contributed reagents; LFM, Conception and design, Analysis and interpretation of data, Drafting and revising the article, Contributed unpublished essential data or reagents; CS, Acquisition of data, Analysis and interpretation of data, Drafting and revising the article; WO, Acquisition of data, Contributed reagents, Drafting and revising the article; DBF, EAB, Conception and design, Contributed reagents, Drafting and revising the article; HEG, MAN, Conception and design, Analysis and interpretation of data, Drafting and revising the article

References

Bayer EA, Belaich JP, Shoham Y, Lamed R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* **58**:521–554. doi: 10.1146/annurev.micro.57.030502.091022

Beitz E. 2000. TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **16**:135–139. doi: 10.1093/bioinformatics/16.2.135

Bertz M, Rief M. 2009. Ligand binding mechanics of maltose binding protein. *Journal of Molecular Biology* **393**:1097–1105. doi: 10.1016/j.jmb.2009.08.066

Brás JL, Alves VD, Carvalho AL, Najmudin S, Prates JA, Ferreira LM, Bolam DN, Romão MJ, Gilbert HJ, Fontes CM. 2012. Novel clostridium thermocellum type i cohesin-dockerin complexes reveal a single binding mode. *The Journal of Biological Chemistry* **287**:44394–44405. doi: 10.1074/jbc.M112.407700

Béguin P, Aubert J-P. 1994. The biological degradation of cellulose. *FEMS Microbiology Reviews* **13**:25–58. doi: 10.1111/j.1574-6976.1994.tb00033.x

Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM. 1999. Atomic force microscopy captures length phenotypes in single proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**:11288–11292. doi: 10.1073/pnas.96.20.11288

Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA. 2003. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **100**:13809–13814. doi: 10.1073/pnas.1936124100

Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith N, Bayer EA, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA, Gilbert HJ. 2007. Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America* **104**:3089–3094. doi: 10.1073/pnas.0611173104

Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, Oparka KJ, Christie JM. 2008. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**:20038–20043. doi: 10.1073/pnas.0807551105

Chen C, Cui Z, Xiao Y, Cui Q, Smith SP, Lamed R, Bayer EA, Feng Y, Cui Q, Smith SP, Lamed R, et al. 2014. Revisiting the NMR solution structure of the Cel48S type-I dockerin module from clostridium thermocellum reveals a cohesin-primed conformation. *Journal of Structural Biology* **188**:188–193. doi: 10.1016/j.jsb.2014.09.006

Cook SM, Lang KM, Chynoweth KM, Wigton M, Simmonds RW, Schäffer TE. 2006. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotechnology* **17**: 2135–2145. doi: 10.1088/0957-4484/17/9/010

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190. doi: 10.1101/gr.849004

Currie MA, Adams JJ, Faucher F, Bayer EA, Jia Z, Smith SP. 2012. Scaffoldin conformation and dynamics revealed by a ternary complex from the clostridium thermocellum cellulosome. *The Journal of Biological Chemistry* **287**:26953–26961. doi: 10.1074/jbc.M112.343897

Demain AL, Newcomb M, Wu JHD. 2005. Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews* **69**:124–154. doi: 10.1128/MMBR.69.1.124-154.2005

Engel A, Müller DJ. 2000. Observing single biomolecules at work with the atomic force microscope. *Nature Structural Biology* **7**:715–718. doi: 10.1038/78929

Fontes CM, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual Review of Biochemistry* **79**:655–681. doi: 10.1146/annurev-biochem-091208-085603

Gumpp H, Stahl SW, Strackharn M, Puchner EM, Gaub HE. 2009. Ultrastable combined atomic force and total internal fluorescence microscope. *Review of Scientific Instruments* **80**:063704. doi: 10.1063/1.3148224

Haimovitz R, Barak Y, Morag E, Voronov-Goldman M, Shoham Y, Lamed R, Bayer EA. 2008. Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. *Proteomics* **8**:968–979. doi: 10.1002/pmic.200700486

Hall BA, Sansom MSP. 2009. Coarse-grained MD simulations and protein−protein interactions: the cohesin−dockerin system. *Journal of Chemical Theory and Computation* **5**:2465–2471. doi: 10.1021/ct900140w

Hoffmann T, Tych KM, Hughes ML, Brockwell DJ, Dougan L. 2013. Towards design principles for determining the mechanical stability of proteins. *Physical Chemistry Chemical Physics* **15**:15767–15780. doi: 10.1039/c3cp52142g

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**:33–38. doi: 10.1016/0263-7855(96)00018-5

Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**:90–95. doi: 10.1109/MCSE.2007.55

Hutter JL, Bechhoefer J. 1993. Calibration of atomic-force microscope tips. *Review of Scientific Instruments* **64**: 1868 doi: 10.1063/1.1143970

Jobst MA, Schoeler C, Malinowska K, Nash MA. 2013. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *Journal of Visualized Experiments* **82** doi: 10.3791/50950

Karpol A, Kantorovich L, Demishtein A, Barak Y, Morag E, Lamed R, Bayer EA. 2009. Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction. *Journal of Molecular Recognition* **22**:91–98. doi: 10.1002/jmr.926

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**:363–371. doi: 10.1038/nprot.2009.2

Kufer SK, Puchner EM, Gumpp H, Liedl T, Gaub HE. 2008. Single-molecule cut-and-paste surface assembly. *Science* **319**:594–596. doi: 10.1126/science.1151424

Kufer SK, Strackharn M, Stahl SW, Gumpp H, Puchner EM, Gaub HE. 2009. Optically monitoring the mechanical assembly of single molecules. *Nature Nanotechnology* **4**:45–49. doi: 10.1038/nnano.2008.333

Li H, Cao Y. 2010. Protein mechanics: from single molecules to functional biomaterials. *Accounts of Chemical Research* **43**:1331–1341. doi: 10.1021/ar100057a

Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TW, Chudakov DM. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* **4**:555–557. doi: 10.1038/nmeth1062

Otten M, Ott W, Jobst MA, Milles LF, Verdorfer T, Pippig DA, Nash MA, Gaub HE, Milles LF, Verdorfer T, Pippig DA, et al. 2014. From genes to protein mechanics on a chip. *Nature Methods* **11**:1127–1130. doi: 10.1038/nmeth.3099

Pagès S, Bélaïch A, Bélaïch JP, Morag E, Lamed R, Shoham Y, Bayer EA. 1997. Species-specificity of the cohesin-dockerin interaction between clostridium thermocellum and clostridium cellulolyticum: prediction of specificity determinants of the dockerin domain. *Proteins* **29**:517–527.

Pinheiro BA, Proctor MR, Martinez-Fleites C, Prates JA, Money VA, Davies GJ, Bayer EA, Fontesm CM, Fierobe HP, Gilbert HJ. 2008. The clostridium cellulolyticum dockerin displays a dual binding mode for its cohesin partner. *The Journal of Biological Chemistry* **283**:18422–18430. doi: 10.1074/jbc.M801533200

Pinheiro BA, Gilbert HJ, Sakka K, Sakka K, Fernandes VO, Prates JA, Alves VD, Bolam DN, Ferreira LM, Fontes CM. 2009. Functional insights into the role of novel type I cohesin and dockerin domains from clostridium thermocellum. *The Biochemical Journal* **424**:375–384. doi: 10.1042/BJ20091152

Puchner EM, Franzen G, Gautel M, Gaub HE. 2008. Comparing proteins by their unfolding pattern. *Biophysical Journal* **95**:426–434. doi: 10.1529/biophysj.108.129999

Roberts E, Eargle J, Wright D, Luthey-Schulten Z. 2006. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**:382 doi: 10.1186/1471-2105-7-382

Sakka K, Sugihara Y, Jindou S, Sakka M, Inagaki M, Sakka K, Kimura T. 2011. Analysis of cohesin-dockerin interactions using mutant dockerin proteins. *FEMS Microbiology Letters* **314**:75–80. doi: 10.1111/j.1574-6968.2010.02146.x

Salama-Alber O, Jobby MK, Chitayat S, Smith SP, White BA, Shimon LJ, Lamed R, Frolow F, Bayer EA. 2013. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *The Journal of Biological Chemistry* **288**:16827–16838. doi: 10.1074/jbc.M113.466672

Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, Ott W, Fried DB, Bayer EA, Schulten K, Gaub HE, Nash MA. 2014. Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications* **5**:5635–35. doi: 10.1038/ncomms6635

Schoeler C, Bernardi RC, Malinowska KH, Durner E, Ott W, Bayer EA, Schulten K, Nash MA, Gaub HE. 2015. Mapping mechanical force propagation through biomolecular complexes. *Nano Letters* **15**:7370–7376. doi: 10.1021/acs.nanolett.5b02727

Schwarz WH. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology* **56**:634–649. doi: 10.1007/s002530100710

Slutzki M, Jobby MK, Chitayat S, Karpol A, Dassa B, Barak Y, Lamed R, Smith SP, Bayer EA, Barak Y, et al. 2013. Intramolecular clasp of the cellulosomal ruminococcus flavefaciens ScaA dockerin module confers structural stability. *FEBS Open Bio* **3**:398–405. doi: 10.1016/j.fob.2013.09.006

Smith SP, Bayer EA. 2013. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Current Opinion in Structural Biology* **23**:686–694. doi: 10.1016/j.sbi.2013.09.002

Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA, Gaub HE. 2012. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America* **109**:20431–20436. doi: 10.1073/pnas.1211929109

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**:207–234. doi: 10.1016/j.pep.2005.01.016

Valbuena A, Oroz J, Hervas R, Vera AM, Rodriguez D, Menendez M, Sulkowska JI, Cieplak M, Carrion-Vazquez M. 2009. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences of the United States of America* **106**:13791–13796. doi: 10.1073/pnas.0813093106

Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques* **26**:680–682.

Woodside MT, Block SM. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annual Review of Biophysics* **43**:19–39. doi: 10.1146/annurev-biophys-051013-022754

van der Walt Stefan, Colbert SC, Varoquaux Gael. 2011. The NumPy array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **13**:22–30. doi: 10.1109/MCSE.2011.37

4.5 Associated Publication P5

Combining in Vitro and in Silico Single Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

by

Tobias Verdorfer, Rafael C. Bernardi, Aylin Meinhold, Wolfgang Ott,

Zaida Luthey-Schulten, Michael A. Nash, and Hermann E. Gaub

published in

Journal of the American Chemical Society, 139 (49), 17841–17852, (2017)

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

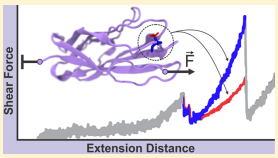
Combining *in Vitro* and *in Silico* Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

Tobias Verdorfer,[†] Rafael C. Bernardi,[‡] Aylin Meinhold,[†] Wolfgang Ott,[†] Zaida Luthey-Schulten,[‡,§] Michael A. Nash,[*,∥,⊥] and Hermann E. Gaub[†]

[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[‡]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[§]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[∥]Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[⊥]Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland

Supporting Information

ABSTRACT: Cellulosomes are polyprotein machineries that efficiently degrade cellulosic material. Crucial to their function are scaffolds consisting of highly homologous cohesin domains, which serve a dual role by coordinating a multiplicity of enzymes as well as anchoring the microbe to its substrate. Here we combined two approaches to elucidate the mechanical properties of the main scaffold ScaA of *Acetivibrio cellulolyticus*. A newly developed parallelized one-pot *in vitro* transcription−translation and protein pull-down protocol enabled high-throughput atomic force microscopy (AFM)-based single-molecule force spectroscopy (SMFS) measurements of all cohesins from ScaA with a single cantilever, thus promising improved relative force comparability. Albeit very similar in sequence, the hanging cohesins showed considerably lower unfolding forces than the bridging cohesins, which are subjected to force when the microbe is anchored to its substrate. Additionally, all-atom steered molecular dynamics (SMD) simulations on homology models offered insight into the process of cohesin unfolding under force. Based on the differences among the individual force propagation pathways and their associated correlation communities, we designed mutants to tune the mechanical stability of the weakest hanging cohesin. The proposed mutants were tested in a second high-throughput AFM SMFS experiment revealing that in one case a single alanine to glycine point mutation suffices to more than double the mechanical stability. In summary, we have successfully characterized the force induced unfolding behavior of all cohesins from the scaffoldin ScaA, as well as revealed how small changes in sequence can have large effects on force resilience in cohesin domains. Our strategy provides an efficient way to test and improve the mechanical integrity of protein domains in general.



INTRODUCTION

Multidomain protein scaffolds organize cellulolytic enzymes and provide adhesion between the host cell and its substrate. In cellulosomes, these so-called scaffoldins utilize various orthogonal high-affinity receptor−ligand interactions between cohesins and dockerins to anchor themselves to the cell's exterior membrane and to coordinate a broad arsenal of cellulolytic enzymes.[1−3] Cellulosomes are extracellular polyprotein complexes produced by many microorganisms for the efficient degradation of cellulose and hemicellulose, two of nature's most abundant polymers. In addition to multiple repeats of cohesin domains, scaffoldins may also contain other ancillary domains, such as enzymatic subunits or carbohydrate-binding modules (CBMs).[4] Cellulosomes represent an advantageous strategy compared to the secretion of freely soluble enzymes for cellulose degradation, because they achieve both robust adhesion to the substrate and synergistic and cooperative interplay between the enzymes. This complex synergy is based on the spatial organization and enhanced local concentration. Due to their remarkable ability to achieve these complex tasks, cellulosomes have become a prime instructive example of molecular systems offering modularity, self-assembly, and highly efficient enzymatic catalysis.[5] Aside from the unique role they play in the degradation of cellulolytic material, cellulosomal scaffoldins are especially interesting as building blocks in a biotechnological toolbox. Cohesin−dockerin interactions with orthogonal specificities can be used to post-translationally incorporate selected enzymes or other auxiliary domains in specific locations by self-assembly within rationally designed multicomponent complexes.[6−8]

In this study, we focused on a scaffoldin of the cellulosome-producing organism *Acetivibrio cellulolyticus*, a gram-negative, anaerobic bacterium found in mechanically demanding environments, such as sewage sludge[9] or the bovine rumen.[10] Although *A. cellulolyticus* expresses an exceptionally branched and diverse

Received: July 28, 2017
Published: October 23, 2017

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Journal of the American Chemical Society Article

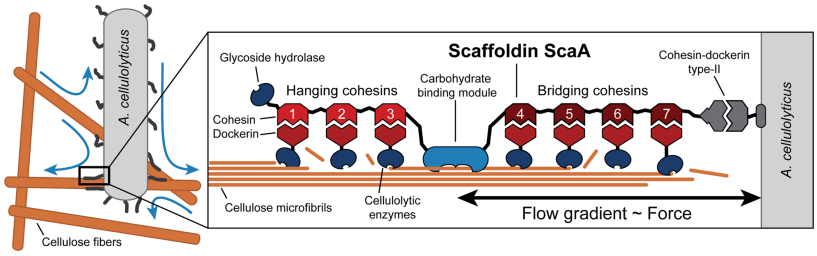


Figure 1. Simplified schematics of the cellulosome of *A. cellulolyticus*. The scaffoldin ScaA binds up to 7 cellulolytic enzymes via cohesin−dockerin interactions and attaches the multienzyme complex to the cellulose microfibril through a CBM module, ensuring close proximity of the enzymes and substrate. Simultaneously, ScaA anchors the assembled cellulosome to the cell wall of the microbe via a type-II cohesin−dockerin complex. The bridging cohesins 4−7 in particular have to withstand the forces between the cellulose fibril and microbe caused by flow gradients in the environment.

cellulosome,[11] the assembly of the majority of its cellulolytic enzymes relies mainly on a single scaffoldin (ScaA)[12] (Figure 1). Simultaneously, ScaA has the essential function of targeting specific substrates and anchoring the cell to the cellulose fibril via its single CBM. ScaA therefore serves a dual purpose for the organism by incorporating catalytic modules into the cellulosome complex and by adhering the bacterium to the cellulosic substrate, making it the key player in *A. cellulolyticus* cellulosome.

ScaA consists of an N-terminal glycoside hydrolase domain, seven type-I cohesins, a CBM located between cohesins 3 and 4, and a C-terminal type-II dockerin domain. Cohesins 4 through 7 are located between the CBM and the anchoring type-II Doc and form the so-called "bridging" region of the scaffoldin, while cohesins 1 through 3 are located exterior from the CBM and form the "hanging" region of this scaffoldin. It had been hypothesized that the cohesins of the bridging region will be subjected to higher mechanical stress compared to the hanging cohesins because the CBM anchors the cell to the cellulose substrate and the domains located within the hanging region are not expected to be mechanically stressed *in vivo*.[13] The organism may have adapted the bridging cohesins to these conditions, and therefore, they should be able to withstand higher mechanical stress in order to maintain a folded structure.

This hypothesis of bridging versus hanging cohesins within cellulosomal scaffoldins has previously been successfully tested in part for the CipA scaffoldin of *Clostridium thermocellum*. These prior results showed that cohesins from the bridging region withstood higher unfolding forces compared to those from the hanging region.[13] This behavior has been attributed to slight differences in the stability of the mechanical clamp motifs, which are structural elements formed by backbone hydrogen bonds between parallel β-strands of the N- and C-terminal ends of the cohesin protein domains.

Since the mechanobiology of cellulosomes is key to their function, we investigated the mechanostability of cohesins of the ScaA scaffoldin as well as several derived mutants at the single molecular level. The high sequence similarity between the selected cohesins raises the question of how differently these cohesins withstand mechanical stress. It is known that small variations in cohesin primary sequences are responsible for determining the specificity of interactions with their dockerin binding partners,[14,15] but very little is known about how sequence variations affect mechanical stability. To address these points, we performed automated atomic force microscope (AFM)-based single molecule force spectroscopy (SMFS) experiments, a technique that has been established as a robust standard approach to investigate the mechanics of individual molecules.[16−22] The fast dynamics and particularly the large force range of AFM-SMFS made it a prime choice for our investigations. To ensure improved relative force comparability and high experimental throughput, we developed a parallelized AFM sample preparation method, which utilizes a one-step protein expression and surface immobilization protocol which is a simplified and easy-to-use version of work previously introduced by our group.[23] At the same time, we carried out steered molecular dynamics (SMD) simulations on structural protein models derived from a homology modeling strategy, which improved our understanding of the cohesin mechanostability at the submolecular level. By employing cross-correlation based network analysis on simulation trajectories, we identified regions involved in structural stability outside the mechanical clamp motif and proposed mutations to pin down single crucial amino acids involved in fold strength. This so-called dynamical network analysis has been successfully employed before to investigate force propagation pathways in cohesin−dockerin complexes[24] and in filamins.[25] We tested the proposed mutants experimentally to verify the influence of the proposed alterations on mechanical stability. Employing this combined computational/experimental approach, we were able to predict and verify, among others, a single point mutation outside of the mechanical clamp motif of cohesin 1, which increased its mechanical stability 2.6-fold. The ability to predict such a remarkable difference in mechanostability reveals the potential of our combined approach to characterize and manipulate the mechanical properties of protein domains.

RESULTS

AFM-Based SMFS. As specific pulling handles throughout this study we used a type-III cohesin−dockerin complex (Coh3:XDoc3) from *Ruminococcus flavefaciens*, which ensured a high yield of single-molecule interaction curves because of its high specificity and long-term stability. This pulling handle enabled the unfolding of mechanically rigid cohesins due to its high receptor−ligand rupture force of more than 600 pN.[26] It is important to note here that the Coh3:XDoc3 interaction, which

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Journal of the American Chemical Society Article

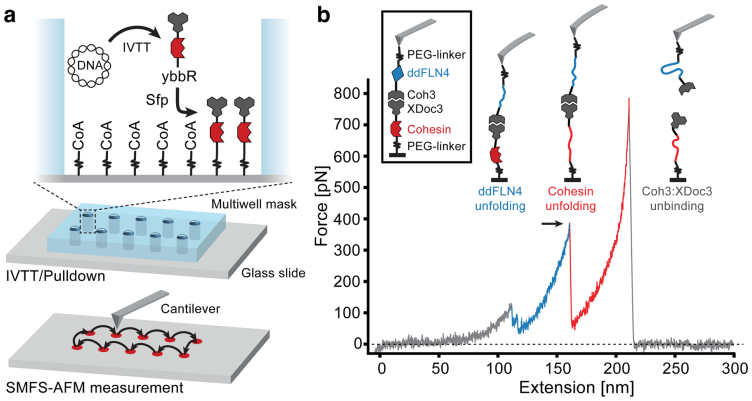


Figure 2. Schematics of the experimental design and exemplary force curve. (a) A multiwell mask is attached to a glass slide, and the surface is functionalized with PEG−Coenzyme A. In a one-pot reaction, an IVTT-kit expresses the proteins of interest containing a ybbR-tag at their N-termini, and an XDoc3 domain at their C-termini. Sfp catalyzes a reaction to bind the constructs site-specifically and covalently to the PEG−CoA spacers. After removal of the multiwell mask, the different constructs are probed by the same AFM tip in an automated fashion. (b) Exemplary force−distance trace showing the unfolding of ybbR−cohesin−XDoc3 immobilized on the surface and the Coh3−ddFLN4−ybbR bound to the cantilever, when the cantilever is retracted from the surface with constant speed. A typical curve shows a two-step unfolding and subsequent stretching of the ddFLN4 fingerprint domain (blue), followed by the unfolding (indicated by an arrow) and stretching of the cohesin domain under investigation (red) and a final rupture of the Coh3:XDoc3 complex.

we employed as a specific pulling handle, is orthogonal to the ScaA cohesin domains that were unfolded under force and thus does not interact with them. ScaA cohesin domains from the bridging or hanging region were cloned in frame with XDoc3 and synthesized using cell-free expression (see below). The cohesins of interest therefore carried the XDoc3 domain at their C-termini, which was able to bind to the Coh3 on the cantilever. The cantilever carried covalently and site-specifically immobilized Coh3 domains, which were recombinantly expressed as fusion proteins with the fourth immunoglobulin rod filamin domain from *Dictyostelium discoideum* (ddFLN4).[27,28] We used the ddFLN4 as fingerprint domain within the molecular chain to reliably rule out data traces with unspecific or multiple interactions. The fingerprint domain unfolds at relatively low forces (∼100 pN) and its contour length increment (∼34 nm) differs significantly from the expected contour length increments of the cohesins (∼45 nm). Furthermore, based on our experience we knew that ddFLN4 reliably refolds following mechanical unfolding when bound to the cantilever.

As calibration errors of up to 15% come with cantilever-based force measurements,[29] the precision and therefore the comparability of individual SMFS experiments is limited by the accuracy of the calibration of the cantilever spring constant. This limitation can be circumvented by presenting a set of molecules in a covalently linked microarray format on a single glass slide in predetermined positions.[23] Taking advantage of this setup, the proteins of interest can all be measured with a single cantilever in a high-throughput fashion, and large sets of molecular constructs can be compared in a single AFM measurement session. Although the systematic error of cantilever calibration will still be up to 15%, the relative stability of the proteins can be assessed with extremely high precision, allowing us to detect differences in stability well below 15%.

We have previously reported the use of microfluidics in combination with an intricate multistep *in vitro* on-chip protein expression and covalent surface attachment protocol.[23] Here we were able to reduce the complexity of the setup significantly, gaining throughput as well as flexibility. We developed a new sample preparation protocol, which is schematized in Figure 2a. Briefly, a silicone mask was placed on a glass slide to form microwells, and the resulting spots were covalently functionalized with PEG−Coenzyme A. The individual wells were filled with an *in vitro* transcription/translation (IVTT) system, along with the plasmid DNA encoding for the fusion proteins and phosphopantetheinyl transferase (Sfp).[30] In a single incubation step, this mixture resulted in cell-free protein synthesis, and simultaneous covalent ligation of the protein library onto the surface through a ybbR-tag[30] at the N-terminus. The mask was removed from the glass slide and the spatially separated protein spots were probed in series using a single functionalized cantilever. The combination of a site-specific N-terminal enzyme-mediated immobilization strategy, and a specific C-terminal pulling handle ensured that only fully expressed constructs were probed by the AFM. We recorded 2000 pulling cycles per protein spot while continuously cycling through the array in an automated fashion. No further user interaction was required after the start of the experiment, which allowed long-term multiday measurements to build up large statistics for each construct.

A typical force versus distance trace is shown in Figure 2b. The cantilever approached the surface, and Coh3 bound to XDoc3. Upon retraction of the cantilever with constant speed, the polypeptide chain stretched until the ddFLN4 fingerprint unfolded in a distinct two-step pattern, followed by the

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

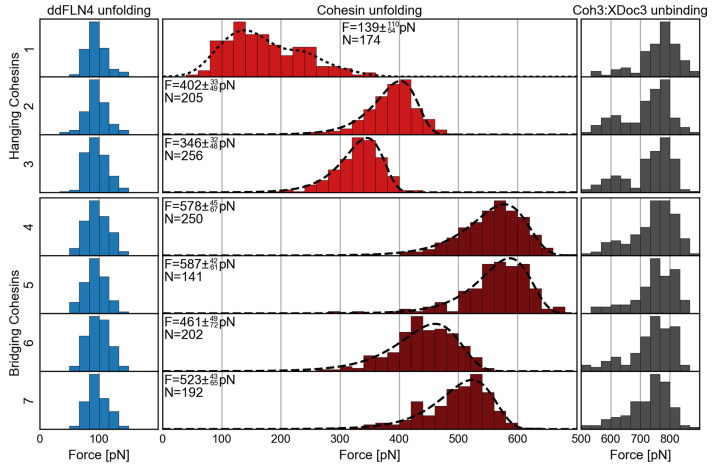


Figure 3. Unfolding and rupture force histograms of wild-type ScaA cohesins. Histograms showing the unfolding and unbinding forces of the fingerprint domain ddFLN4 (blue), the hanging (light red) and bridging (dark red) cohesins of ScaA, and the Coh3:XDoc3 receptor−ligand pulling handle (gray). The force distributions of the ddFLN4 fingerprint and the Coh3:XDoc3 handle are independent of the measured ScaA cohesin construct, which allows for improved relative force comparability of the ScaA cohesins. Unfolding force distributions of cohesins 2−7 were fitted following the Bell−Evans model (dashed lines). A kernel density estimation (KDE) was used to obtain the most probable unfolding force (±fwhm) of cohesin 1 (dotted line). All data were recorded using a single cantilever with a spring constant of 225 pN/nm at a retraction speed of 1600 nm/s during a 24 h automated SMFS experiment.

unfolding of the respective cohesin under investigation. Finally the Coh3:XDoc3 bond ruptured typically at forces of ∼780 pN at force loading rates of ∼10^5 to 10^6 pN/s. The heights of the various peaks are plotted in the histograms of Figures 3 and 5. Only force curves displaying a distinct combination of contour length increments (ddFLN4 ∼ 34 nm,[27,28] cohesin ∼ 45 nm) were included in our analysis. Alignments of all relative contour length increments from all pulling experiments can be found in the Supporting Information (Supplemental Figure S1).

AFM-Based SMFS on ScaA's Wild-Type Cohesins. To investigate the mechanical stability of *A. cellulolyticus*' scaffoldin ScaA, seven distinct surface spots (one for each cohesin) were prepared using the one-step expression/immobilization reaction described above. All data were collected using a single cantilever. Figure 3 shows the resulting unfolding and rupture force distributions. The outer histograms show very similar force distributions of the fingerprint domain ddFLN4 and of the pulling handle Coh3:XDoc3 independent of the measured molecular construct, which agreed with previously reported literature values.[26,27] Coh3:XDoc3 showed a sharp peak at ∼780 pN and a minor shoulder at ∼600 pN, both of which are known features of this molecular complex.[26] The remarkable consistency in force among different molecular constructs indicated low force drift during the course of the experiment and confirmed our ability to achieve precise relative comparability of the unfolding forces of the cohesins. Following work done by the Schulten[31] group and Evans and Ritchie[32] in 1997, the cohesin unfolding force distributions can be fitted using a two-state model (from here on called the Bell−Evans model).

With the exception of cohesin 1, all unfolding force distributions could be fitted using the Bell−Evans model, indicating a stable fold and a single barrier unfolding pathway dominating the unfolding process, and most probable rupture forces were obtained. The measured unfolding forces of cohesin 1 resulted in a more complex force distribution. Individual AFM force−distance traces gave no hints of any peculiarities in comparison to the unfolding traces of the other cohesins. We hypothesize that cohesin 1 exhibits more than one distinct fold or has several multibarrier unfolding pathways that precluded its unfolding force distribution to be fitted using a simple two-state model. In this case, we used kernel density estimation (KDE) as a means to smooth the unfolding force histogram and obtain the most probable unfolding force. As originally hypothesized, the three hanging cohesins showed in fact a considerably lower most probable unfolding force (for full width at half-maximum (fwhm) errors, see Figure 3) (cohesin 1, 139 pN; cohesin 2, 402 pN; cohesin 3, 346 pN) compared to the four bridging cohesins (cohesin 4, 578 pN; cohesin 5, 587 pN; cohesin 6, 461 pN; cohesin 7, 523 pN).

Cohesin Homology Models. Since structural data were not available for any of the ScaA cohesins, a homology modeling strategy was adopted,[33] employing Modeller 9.17[34] to obtain structural models for all the cohesins investigated here (Figure 4b). Using BLAST,[35] we obtained homologous cohesin structures (PDB IDs 1G1K, 4DH2, 2VN6, and 4UMS) within the Protein Data Bank[36] (PDB). These structures were then used as templates to derive the homology models that were further refined with molecular dynamics (MD) simulations. Equilibration for 100 ns was performed using NAMD[37]

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Journal of the American Chemical Society Article

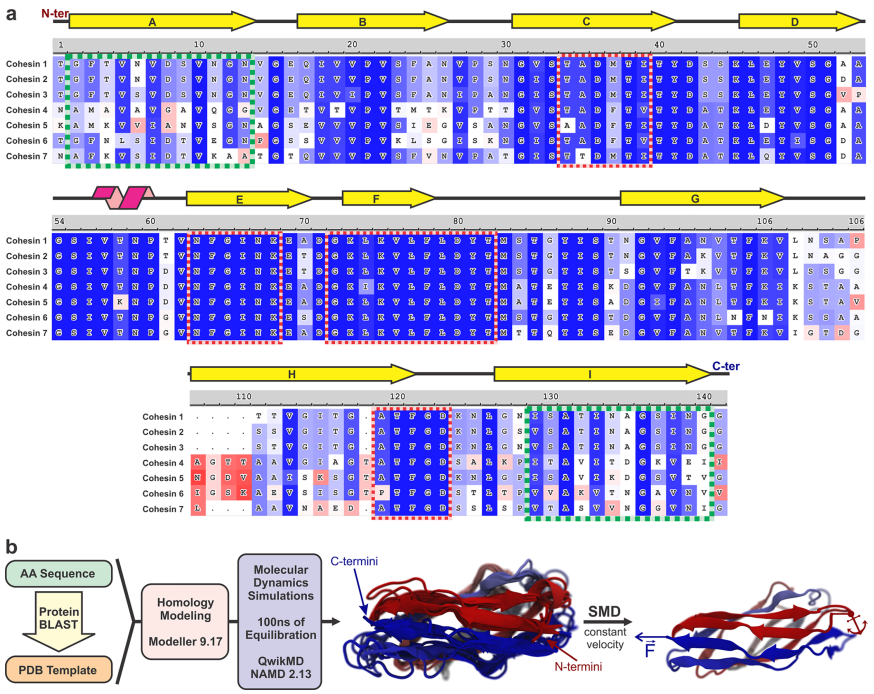


Figure 4. Sequence alignment and modeling workflow. (a) Sequence alignment of all cohesins of ScaA reveals high overall conservation. The amino acids thought to be primarily involved in mechanical stability[13] are represented by green boxes. Regions primarily involved in dockerin recognition and binding are represented in red boxes. The background colors of the letters represent BLOSUM 70 sequence alignment score, from high (dark blue) to low (red). A high resolution version can be found in Supplemental Figure S6. (b) Schematic representation of molecular modeling and dynamics protocol. Homologous protein structures were obtained by running the sequences of the cohesins against the Protein Data Bank using BLAST. Most highly identical structures with high sequence coverage were then used as templates to obtain structural models using Modeller 9.17. Models were properly solvated employing QwikMD and equilibrated for 100 ns of unbiased molecular dynamics simulations using NAMD. A very similar fold was observed for all seven cohesins, here colored ranging from red to blue based on its residue index number. SMD simulations were performed by holding the N-termini and pulling the C-termini with constant velocity.

through its QwikMD interface[38] (see Supplemental Figures S2, S3, and S4). Comparison of the aligned structures (Supplemental Figure S5A) reveals that all seven cohesins of ScaA show a similar general fold. Even though the range of identity between our model cohesins and the best available structural templates was between 33% and 45%, cohesins always present a very similar fold, helping the structure prediction.

SMD Simulations and Network Analyses. To evaluate the behavior of ScaA's cohesins under force, we performed SMD simulations[31,39] using NAMD and QwikMD, where the N-termini of the constructs were fixed and the C-termini were pulled with constant velocity. Employing four different pulling speeds, we investigated first all the unfolding steps in long (on the order of microseconds) SMD simulations. For all cohesins, the first part to unfold with highest peak force is the C-terminal region with β-strand I, followed by β-strand H losing its structure (see Supplemental Figure S7). Next, in the N-terminal region, both β-strands A and B lose their structure almost at the same time in most simulations. β-strands C−G finally lose structure under relatively low force (see representative unfolding analysis in Supplemental Figure S8). The force necessary to break any of the cohesin folds was observed to be equal to the force required to unfold the C-terminal segment of these cohesins. Therefore, to obtain statistically relevant mechanical stability behavior, we performed many 20 ns long simulation replicas (at least 25 replicas per construct). Figure 5a shows the peak unfolding forces, as well as the sequence identity between the cohesins of ScaA and their respective best available structural template. With few exceptions, the simulation replicas revealed qualitatively the same general trend of the unfolding force peak distributions as measured by SMFS (Figure 3). The absolute forces however are shifted in the simulations toward higher values (Supplemental Figure S9), which was to be expected due to

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Journal of the American Chemical Society Article

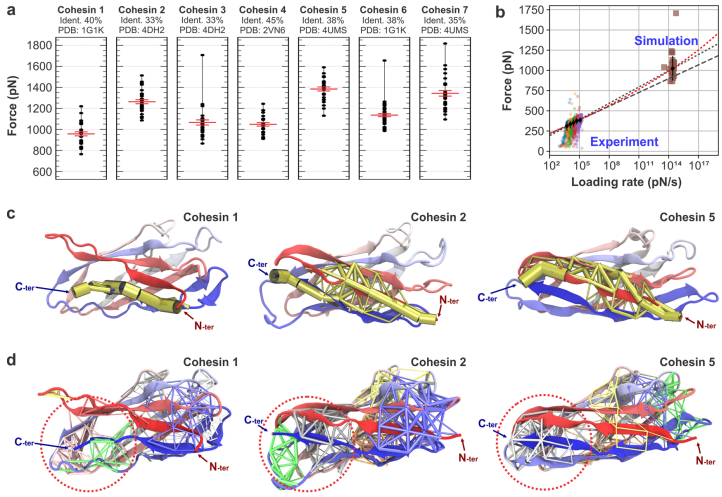


Figure 5. SMD peak force results, dynamic force spectrum, force propagation pathways, and community analysis. (a) SMD peak unfolding forces from each simulation replica (black dots) and average peak force per cohesin (red line ± SEM) for the different natural cohesins of ScaA. Sample force versus distance profiles are found in Supplemental Figure S10. Statistical significance between the unfolding forces of all cohesins can be found in Supplemental Figure S11. Also shown is the sequence identity to the respective PDB homology modeling template. (b) Dynamic force spectrum for unfolding events of cohesin 3. Varicolored points represent rupture force/loading rate data from an experiment with 5 different pulling speeds. Brown squares represent rupture force/loading rate data from SMD simulations. Black points represent the most probable rupture force/loading rate of each pulling speed obtained from kernel density estimates. Error bars represent the fwhm. Gray lines represent least-squares fits of the Bell–Evans model to the experimental and to both the experimental and the simulation data with fitting parameters (±SD) $\Delta x = 0.17 \pm 0.012$ nm, $k_0 = (6.7 \pm 6.3) \times 10^{-4}$ s^{-1} and $\Delta x = 0.14 \pm 0.0015$ nm, $k_0 = 4.9 \times 10^{-3} \pm 8.9 \times 10^{-4}$ s^{-1}, respectively. The red dotted line represents a least-squares fit of the DHS model to the combined experimental and the simulation data with fitting parameters (±SD) of $\Delta x = 0.19 \pm 0.024$ nm, $k_0 = (1.4 \pm 2) \times 10^{-4}$ s^{-1}, and $\Delta G = 60 \pm 13\ k_BT$. For detailed representation of experimental data, see Supplemental Figure S12. (c) Force propagation pathways through selected cohesins calculated using Pearson correlation (yellow tubes). The thickness of the pathway edges represents the normalized probability of force propagating through the particular edge. (d) Network-based community analysis in selected cohesins showing regions with high internal correlation during pulling simulations calculated using generalized correlation. Communities are colored individually, and thick connections correspond to high correlation.

much higher force loading rates of the simulations compared to the AFM experiments.

A force offset between MD simulations and SMFS experiments can easily be understood in view of the Bell–Evans model, which predicts a linear dependence of the unfolding force from the logarithm of the force loading rate. To corroborate this assumption, we varied the pulling rate in the SMFS experiments of cohesin 3 and plotted the measured unfolding forces together with the corresponding MD simulations in Figure 5b. The dashed line represents the best fit to the experimental data extrapolated to the MD time domain; the dotted line fits both the experimental and the simulation data. The resulting distances to the transition state of 0.14 ± 0.012 nm and 0.17 ± 0.0015 nm agree very well with literature values of comparable cohesin protein structures.[13,15,26] It is important to note that the slope in the dynamic force spectrum can change with increasing pulling speeds, resulting in a nonlinear upturn at higher pulling velocities as shown by Rico et al.[18] This effect is caused by a shift from a stochastic to a deterministic unfolding regime. In the former, the unfolding process is governed by spontaneous, thermal unfolding under a given force, while in the latter, the high pulling velocities leave the protein insufficient time to sample its energy landscape. As described in the Dudko, Hummer, and Szabo model (DHS model), the regime transition can happen at different loading rates and is characterized by the critical force, $F_c = 2\Delta G/\Delta x$.[40] The position of the transition from stochastic to deterministic regime therefore strongly depends on the general mechanical stability of the system under investigation. Fitting the DHS model to the data in Figure 5b resulted in a critical force $F_c \approx$ 2500 pN, suggesting that our SMD simulations were carried out at loading rates where unfolding is still dominated by stochastic fluctuations. This indicates that the SMD simulations provide an accurate description of the unfolding process of the system in this study.

The aforementioned results motivated a detailed analysis of the molecular structures and interactions, which could give rise to the particular properties of the different cohesins. Using Pearson cross-correlation-based force propagation analysis, a recently introduced protocol for the analysis of load distributions in molecular complexes,[24] we calculated the

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841–17852

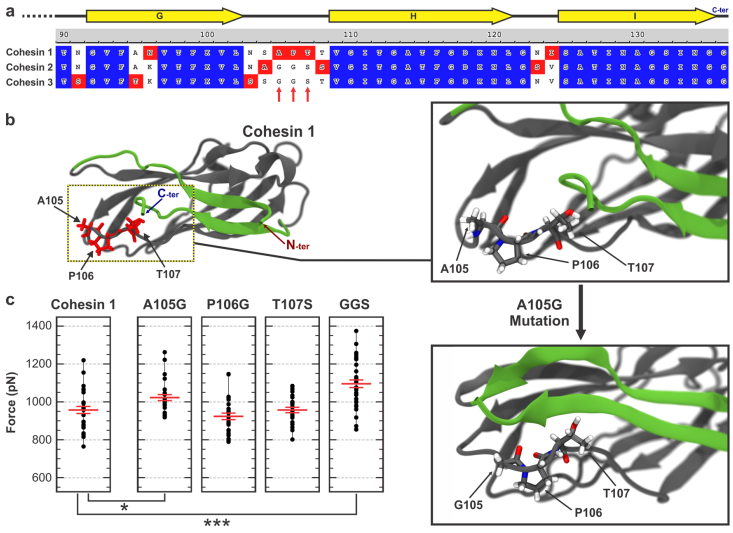


Figure 6. Mutant design and SMD peak force results. (a) Sequence alignment of the C-terminal end of the hanging cohesins 1−3. From the community analysis, we deduced that the low force resilience of cohesin 1 originated from the loss of sequence identity in the flexible area around amino acids 103−108. Based on comparison with cohesins 2 and 3, we designed three point mutations (A105G, P106G, and T107S) and a triple mutant (A105G P106G T107S (referred to as mutant "GGS")). A high resolution version can be found in Supplemental Figure S15. (b) Left, homology model of cohesin 1 showing the β-sheets involved in the mechanical clamp motif in green and unique amino acids within the group of hanging cohesins in red. The linker region around amino acids A105, P106, and T107 was observed to be most flexible in the SMD simulations and was therefore suggested for mutation studies. Right, comparison of the homology models of cohesin 1 and its mutant A105G with molecular representation of the aforementioned region of interest. A seemingly small point mutation from an alanine to a glycine (from a methyl to a proton side chain) at position 105 changed the fold of the protein in this region significantly, resulting in much closer β-strands in the mechanical clamp motif. (c) SMD peak force from each simulation replica (black dot) and average peak force per cohesin (red line ± SEM). The significantly increased peak forces of the mutants A105G and GGS suggest an increased stability compared to wild-type cohesin 1, while mutants P106G and T107S showed no significant change in average peak unfolding force. The single A105G mutation was able to recover forces in the same range of cohesin 3 (see Figure 5a). Statistical significance between the unfolding forces of all cohesins can be found in Supplemental Figure S16.

suboptimal force propagation pathways, revealing that force propagates mainly through β-strands A, B, and I (see Supplemental Figure S13). Figure 5c shows the force propagation pathways through cohesin 1 and the two strongest cohesins within their respective group, cohesins 2 and 5. Cohesin 1 predominantly showed a single path between the N- and C-terminal β-sheets that carried the entire mechanical load, suggesting a badly formed mechanical clamp motif. The other cohesins show multiple possible force pathways suggesting a better distribution of force propagating from the N- to the C-terminus through a multitude of backbone hydrogen bonds (see Supplemental Figure S13).

In addition, we investigated the communities formed in the systems by employing dynamic network analysis[41] and generalized correlation[42] (see Supplemental Figure S14). These communities correspond to sets of residues that move in concert with each other and can be used to investigate regions that are generally more strongly connected during pulling simulations. Figure 5d shows the individual communities of cohesins 1, 2, and 5 in different colors, where thicker connections between the amino acids correspond to higher correlation between them. While cohesins 2 and 5 showed pronounced communities connecting the C-terminus to its surroundings (Figure 5d, red dashed circle region), cohesin 1 showed weak communities in this area of the protein, suggesting high and uncorrelated flexibility and therefore loose intraprotein contacts between the N- and C-terminal β-sheet. Particularly, the area between β-sheets G and H turned out to be most flexible in the case of cohesin 1. The same region of cohesin 1 was observed to be highly flexible also during the 100 ns MD equilibration, and was not as flexible in any other cohesin investigated here. Taken together, our results suggested that this is a critical region responsible for cohesin 1 lower stability under mechanical force.

Mutant Design and SMD Simulations. Since the region between β-sheets G and H in cohesin 1 was found to be the most flexible during the equilibrium MD and the analysis of cross-correlation communities suggested that weak communities in this same area could be responsible for the badly formed mechanical clamp motif of cohesin 1, we aligned the sequences of all hanging cohesins in this region (Figure 6a). Despite the high overall sequence similarity of ∼85% among *A. cellulolyticus*'s hanging cohesins, major differences exist between the weaker cohesin 1 and the stronger cohesins 2 and 3 in the

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Journal of the American Chemical Society Article

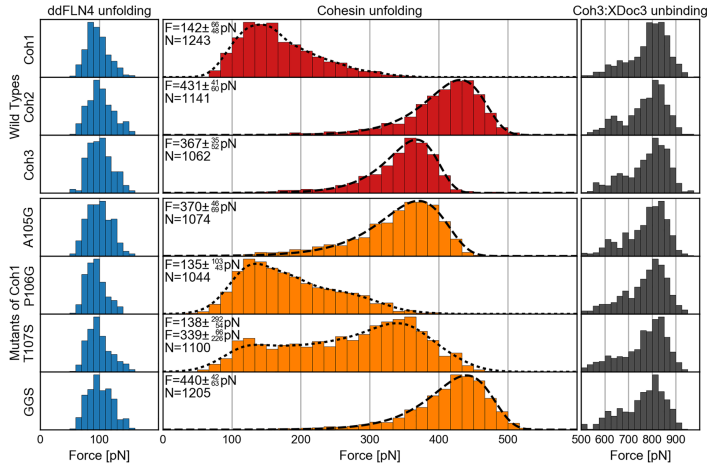


Figure 7. Unfolding and rupture force histogram of the hanging cohesins and mutants of cohesin 1. Histograms showing the unfolding and unbinding forces of the fingerprint domain ddFLN4, the wild-type hanging cohesins 1−3 (red), mutants of cohesin 1 (orange), and the Coh3:XDoc3 receptor ligand binding handle (gray). The force distributions of the ddFLN4 fingerprint and the Coh3:XDoc3 handle are independent of the measured ScaA cohesin construct. Similar to Figure 3, force histograms were fitted following a Bell−Evans model where possible (cohesin 2, cohesin 3, A105G, and GGS; dashed line). A KDE was used to find the most probable rupture forces (±fwhm) in all other cases (cohesin 1, P106G, and T107S; dotted line). All data were recorded using a single cantilever with a spring constant of 163 pN/nm at a retraction speed of 1600 nm/s during a 72 h automated SMFS experiment.

region between β-strands G and H, more precisely from amino acid 103 to 108. Considering the results of the SMD simulations, force propagation, community analysis, and the sequence comparison with cohesin 2 and 3, we proposed four mutants of cohesin 1, namely, A105G, P106G, and T107S, and a triple mutant that we refer to here as GGS (A105G, P106G, and T107S). To investigate if these mutants would show a higher force resilience compared to their wild-type counterparts, we followed the same modeling and simulation steps as previously discussed for the wild-type cohesins.

After the 100 ns MD equilibration, mutated cohesins showed significant structural differences compared to the wild-type cohesin 1. A single alanine to glycine mutation (A105G), for example, already stabilized the nearby regions of the protein, resulting in a flawlessly folded β-stranded C-terminus, as shown in Figure 6b. SMD simulations, as shown in Figure 6c, revealed that A105G and GGS displayed a significantly higher unfolding force than cohesin 1, with A105G showing a mean unfolding force similar to that of cohesin 3. Analysis of the force propagation profiles and communities of the A105G mutant revealed a behavior that resembles one of the stronger cohesins, as shown in Supplemental Figures S13 and S14.

AFM-Based SMFS on Hanging Cohesins and Mutants of Cohesin 1. In order to test the predictions from the SMD simulations and to identify the amino acids responsible for the low force resilience of cohesin 1, we prepared a second set of AFM-based SMFS experiments similar to the ones described in Figure 3. We compared the mechanical stability of the proposed mutants A105G, P106G, T107S, and the triple mutant GGS with the three hanging cohesins, again using a single cantilever to ensure improved relative force comparability (Figure 7). Similar to Figure 3, force distributions not following the Bell−Evans model (cohesin 1, P106G, and T107S) were smoothed using a KDE to obtain meaningful most probable rupture forces. Most remarkably, mutant A105G showed a dramatic increase in most probable rupture force to 370 pN, making it around 2.6 times stronger than its parent structure cohesin 1 and therefore about as strong as cohesin 3. While the mutant unfolding simulations predicted a slight decrease in mechanical stability for the P106G mutant, the experimentally obtained unfolding force histogram shows no considerable change compared to wild-type. Mutant T107S exhibited a bimodal unfolding force distribution with most probable rupture forces of 138 pN and 339 pN roughly similar to the most probable unfolding forces of wild-type cohesins 1 and 3. A detailed examination of individual unfolding traces from different force regimes showed no distinctive features that could explain its bimodal unfolding force distribution. We can only theorize that this construct might exhibit a combination of strongly differing folded conformations or unfolding pathways. Such behavior was not observed in the simulations. The triple mutant GGS showed a most probable rupture force of 440 pN, making it as strong as cohesin 2.

DISCUSSION

Mechanical Stability of Highly Homologous ScaA Cohesins. The high precision comparison of the mechanical stability of seven homologous cohesin domains from *A. cellulolyticus*' scaffoldin ScaA was enabled by the development of a novel SMFS sample preparation method, where several

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

constructs were produced by cell-free *in vitro* expression and covalent linkage to the surface in parallel in individual microwells. The proteins were probed sequentially with a single cantilever, enabling precise comparison of unfolding force distributions between multiple domains by eliminating relative error in calibration of cantilever spring constant values. All constructs contained a ddFLN4 domain, which served as a molecular fingerprint and allowed clear identification of single-molecule force−distance traces showing specific interactions through their unique contour length increments. This overall approach facilitated high-throughput SMFS of multiple proteins and allowed for fast and automated data analysis.

A clear trend can be seen in the most probable unfolding force of all cohesin domains from ScaA (Figure 3). The three cohesins on the far side of the substrate-anchoring CBM (i.e., hanging cohesins) showed in fact a considerably lower most probable unfolding force compared to the four bridging cohesins, which are exposed to force *in vivo*. This result strongly supports the hypothesis that higher mechanical stability is a repeatable feature of cohesins in the bridging region of cellulosomal scaffoldins, despite the fact that all domains tested show high levels of sequence homology (see Figure 4a).

To elucidate the origins of the large differences in mechanostability of the cohesin domains, we generated structural homology models for all seven cohesins under investigation. The combination of knowledge-based information from structural templates together with modern force fields and molecular dynamics allowed us to employ a real-space structural prediction and refinement strategy to obtain all ScaA cohesin structures.[33] However, the computational prediction of three-dimensional protein structures has its limitations, and the accuracy of the predicted models is strongly dictated by the availability of close structural templates.[43] The range of identity between our model cohesins and the best available structural templates was between 33% and 45%, implying rather poor homologues. To check the fold stability, all structural models were subjected to 100 ns of equilibrium MD, and the final structures were superimposed. The results showed that, even though four different templates were employed, all cohesins generated highly similar structural models (see Figure 4b and Supplemental Figure S5).

The seven structural models were then stretched *in silico* using a constant velocity SMD protocol. It is noteworthy that, with the exception of cohesin 4, the forces of all distributions were shifted by a constant value (±SD) of 782 ± 29 pN (see Supplemental Figure S9). This finding is remarkable if one takes into account the relatively low identity between the modeled systems and their templates (33%−45%). Simulations showed that although cohesin 4 has the highest identity to its template (45%, PDB 2VN6), it might have been a suboptimal choice resulting in a nonideal folding state, as it shows an N-terminal region with wobbly β-strand formation (see Supplemental Figures S4 and S5).

Comparing the force peaks between the simulations and experiments served as a validation for the homology structures. Figure 5b shows a direct comparison between results obtained with AFM SMFS across a range of loading rates from $\sim 10^3$ pN/s and $\sim 10^5$ pN/s, and those obtained from in silico SMFS at $\sim 10^{14}$ pN/s. Fitting the DHS model to the data suggests that the loading rates used in our SMD simulations fall into the stochastic regime. This finding, which bridges 11 orders of magnitude in force loading rate, indicates that the homology models provide an accurate description of the unfolding process, validating the predictive power of both comparative modeling and in silico SMFS. We want to emphasize that this only holds true for remarkably strong proteins like the cohesins investigated here. At similar loading rates of $\sim 10^{14}$ pN/s, weaker systems may be unfolded in the deterministic regime. In this case, a slower pulling velocity would have to be chosen, requiring considerably more computational time.

Investigation of the Low Force Resilience of Cohesin 1. Our simulation results in combination with calculated force propagation pathways and correlation communities suggested that the high flexibility in the region around amino acids 100−110 could be responsible for a badly formed mechanical clamp between the N- and C-terminal β-sheets of cohesin 1. We proposed the aforementioned mutants, A105G, P106G, T107S, and the triple mutant GGS carrying all three mutations, in an attempt to affect the folding and the formation of the mechanical clamp motif and, ideally, improve mechanical stability.

Following the same modeling and equilibration protocol followed previously, we obtained structural models for the mutants. An inspection of these structures, after 100 ns of MD, revealed how the fold can be affected by a single A105G mutation, as shown in Figure 6b. The A105G point mutation resulted in a longer mechanical clamp between the N- and C-terminal β-sheets. SMD simulations predicted an increase in unfolding forces both for mutant A105G and for mutant GGS, the latter being the most promising as shown in Figure 6c. It is noteworthy that the simulations revealed that the single A105G mutation already increases the force necessary to unfold cohesin 1 up to the same levels of the cohesin 3.

We want to emphasize that a strategy of carrying out only a couple of SMD simulations with low pulling velocity might give an incomplete picture of a biomolecular system under shear force. The approach adopted here, of simulating many fast pulling simulations (totaling over 350 independent SMD runs), showed that the force distribution in simulations is as widely distributed as in experiments, and therefore a small *in silico* sampling might reveal differing trends to those observed experimentally. A possibility to sample both with slow pulling and many replicas would be to employ coarse-grained methods, which are less computationally demanding. However, our simulations revealed that seemingly small mutations, like changing a methyl group to a hydrogen, can cause enormous differences in folding and therefore force resilience. With such minor changes in the biomolecule, exploratory studies to design new mutants using coarse-grained molecular dynamics simulations would be hardly reliable, as they would lack atomic detail.

In order to test the predictions from the SMD simulations, we compared the proposed mutants to the wild-type hanging cohesins experimentally (Figure 7). We found that the two promising mutants, A105G and GGS, showed a considerable increase in mechanical stability. Mutant A105G showed an increase of most probable rupture force by nearly 2.6-fold to 370 pN, relative to its wild-type cohesin 1, which unfolded at 142 pN. As predicted by the SMD simulations, this seemingly small change from an alanine to a glycine outside of the mechanical clamp motif influenced the fold of the protein enough to make it as strong as cohesin 3. The triple mutant GGS showed, again as predicted by the SMD simulations, the largest increase in unfolding force to 440 pN, making it as strong as cohesin 2, the strongest cohesin within the group of

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

hanging cohesins, which unfolded at 431 pN. The experimental results confirm the amino acids responsible for the low force resilience of cohesin 1 and the predicted increase in mechanostability of the proposed mutants, thus corroborating the *in silico* approach.

CONCLUSIONS

Multienzyme molecular devices like cellulosomes rely on scaffoldins for the organization of their active constituents. Since these large protein structures can sometimes be subjected to sizable forces, their mechanical stability is a prerequisite for proper and sustained function. This holds particularly true for the scaffoldin ScaA of *A. cellulolyticus*, which anchors the microbe to a cellulose fibril through a CBM in addition to spatially organizing an ensemble of cellulolytic enzymes. Another interesting aspect is that cellulosome scaffoldins are mainly composed of highly similar cohesin domains with very different mechanical properties. Our *in vitro* and *in silico* SMFS studies not only show that the mechanical stability of all cohesins from the scaffoldin ScaA is consistent with the hypothesis proposed by Valbuena et al.[13] that bridging cohesins are mechanically stronger than hanging cohesins but, moreover, how minimal differences in protein sequence can lead to very different behavior under shear force.

We elucidated the surprisingly low unfolding force of cohesin 1, when compared to the other ScaA cohesins. We found that the point mutation A105G increased the mechanical stability of cohesin 1 more than 2-fold when compared to wild-type. The remarkably strong influence on the mechanical stability of cohesin 1 of a single alanine to glycine mutation, which effectively only substitutes a methyl group by a hydrogen atom, raises the question why evolutionary pressure has not favored this mutant, as it comes at virtually no additional cost for the organism. Possibly not all cohesins are supposed to display high mechanical stability, since cellulosomal organisms have already been shown to be able to regulate their gene expression patterns depending on potentially varying substrates.[44,45] Thus, occasionally un- and refolding cohesins would ensure that cellulosomal components can be exchanged in case of changing environmental conditions.

Both approaches, *in silico* and *in vitro*, of our combined approach started from the genetic information coding for the protein, from which the homology models for the former were derived and the samples for the latter were expressed. Given the large number of cellulosome producing microorganisms with sequenced genomes, a wide spectrum of novel combinations, for example, cohesin−dockerin pairs with similar or orthogonal affinities and tunable strengths, may be analyzed, modified, and combined. The fact that cellulosomes are extracellular organelles of microbes that live in largely diverse ambient environments, including the human gut,[46,47] guarantees robustness of its molecular building blocks and their interactions. This is reflected in the extremely high unfolding barriers and rupture forces of its molecular constituents and qualifies them for a large range of potential applications.

As viable candidates for source materials in a rationally designed artificial protein nanomachine, cellulosomes have demonstrated large potential in molecular engineering applications.[1,6,8] The development of recombinant designer cellulosomes using so-called chimeric scaffoldins allowed control over the position of each enzyme in the cellulosomal complex.[7] Synthetic scaffolds containing orthogonal cohesin domains have furthermore been successfully displayed on the surface of yeast cells, allowing dockerin-tagged cellulases to bind and improve ethanol production almost by a factor of 3 compared to free enzymes.[48] In terms of industrial cellulose degradation, the incorporation of mechanically stronger cohesin domains and cohesin−dockerin interactions with higher affinities will make designer cellulosomes more durable and efficient. A better understanding of individual cellulosomal components can improve upcoming designs and lead to more efficient and reliable multienzyme molecular devices. For example, the new-found properties of ScaA recommend this scaffoldin and its cohesin domains to be part of a potential versatile molecular breadboard for the programmed self-assembly of molecular devices with designed properties.

From a technical point of view, we were able to measure seven constructs using a single cantilever in two separate experiments by utilizing a fast and parallelized sample preparation method, while still achieving sufficient statistics (N = 1420 in 24 h and N = 7869 in 72 h). Furthermore, we have shown that even in the absence of crystallized protein structures, SMD simulations, when combined with protein homology modeling, are a powerful tool to investigate the intricate mechanisms governing protein mechanics. Particularly force propagation and community analyses have proven instrumental, not only allowing us to analyze the origins of a particular molecular property, such as the low mechanical stability of cohesin 1, but also opening new means to identify crucial regions for point mutations aiming at locally altering the mechanics of the protein of choice. In summary, our newly developed methods are enabling novel investigations of protein unfolding and rational modification of structural aspects of proteins based on common design principles across different families of proteins well beyond the cellulosome community. Our results demonstrate a strategy that can be applied in fine-tuning mutations that can change the mechanostability of protein domains and also raise further questions about the evolutionary pressures that can result in mechanically stronger or weaker proteins. Considering the vast number of cellulosomal constituents yet to be explored, the combination of techniques presented here can potentially accelerate the probing and design of scaffolding domains, starting from nothing more than their genetic code, presenting new opportunities in molecular engineering and biotechnology.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b07574.

Materials and methods (gene construction, protein purification, AFM sample preparation, one-step *in vitro* expression and protein pulldown, AFM SMFS measurements, AFM SMFS data analysis, structural model determination, and molecular dynamics simulations), Supplementary Figures S1−S16, and protein sequences (PDF)

AUTHOR INFORMATION

Corresponding Author

*michael.nash@unibas.ch

ORCID

Hermann E. Gaub: 0000-0002-4220-6088

Notes

The authors declare no competing financial interest.

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

■ ACKNOWLEDGMENTS

Support for this work was provided by the EU seventh Framework Programme NMP4-SL-2013-604530 (CellulosomePlus), the Nanosystems Initiative Munich, and the ERC Advanced Grant CelluFuel. M.A.N. acknowledges support from Society in Science – The Branco Weiss Fellowship from ETH Zurich, from an ERC Starting Grant (no. 715207), and from the Human Frontier Science Program (grant no. RGY0080/2015). This work was supported by National Institutes of Health (NIH) grant 9P41GM104601, "Center for Macromolecular Modeling and Bioinformatics". R.C.B. is partially supported by the National Science Foundation (NSF) grant MCB-1616590, "Molecular Modeling of Bioenergetic Systems", as well as the Energy Biosciences Institute (Regents of the University of California Berkeley, EBI 231 UCB BP 2014OO4J01). Molecular dynamics simulations used for structure prediction and equilibration made use of Blue Waters supercomputer as part of the Petascale Computational Resource (PRAC) grant "The Computational Microscope", which is supported by the National Science Foundation (award number ACI-1440026). Blue Waters sustained-petascale computing project is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Steered molecular dynamics simulation replicas made use of the Argonne Leadership Computing Facility (ALCF)/Mira supercomputer as part of the DoE ALCC program. This research used resources of the ALCF, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. T.V. thanks Lukas F. Milles for providing the force curve analysis software, as well as E. Durner, M. A. Jobst, and W. Vanderlinden for helpful discussions and Thomas Nicolaus and Angelika Kardinal for laboratory assistance. We thank Edward A. Bayer and Marcelo Melo for helpful discussions.

■ REFERENCES

(1) Bayer, E. A.; Morag, E.; Lamed, R. *Trends Biotechnol.* **1994**, *12* (9), 379.
(2) Bayer, E. a.; Lamed, R.; White, B. a.; Flint, H. J. *Chem. Rec.* **2008**, *8* (6), 364.
(3) Carvalho, A. L.; Dias, F. M. V.; Prates, J. A. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romão, M. J.; Fontes, C. M. G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13809.
(4) Ding, S.-Y.; Bayer, E. A.; Steiner, D.; Shoham, Y.; Lamed, R. *J. Bacteriol.* **1999**, *181* (21), 6720.
(5) Ding, S.-Y.; Xu, Q.; Crowley, M.; Zeng, Y.; Nimlos, M.; Lamed, R.; Bayer, E. A.; Himmel, M. E. *Curr. Opin. Biotechnol.* **2008**, *19* (3), 218.
(6) Xu, Q.; Ding, S.-Y.; Brunecky, R.; Bomble, Y. J.; Himmel, M. E.; Baker, J. O. *Biotechnol. Biofuels* **2013**, *6* (1), 126.
(7) Fierobe, H.-P.; Mingardon, F.; Mechaly, A.; Bélaïch, A.; Rincon, M. T.; Pagès, S.; Lamed, R.; Tardif, C.; Bélaïch, J.-P.; Bayer, E. a. *J. Biol. Chem.* **2005**, *280* (16), 16325.
(8) Gefen, G.; Anbar, M.; Morag, E.; Lamed, R.; Bayer, E. a. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (26), 10298.
(9) Khan, A. W. *Microbiology* **1980**, *121* (2), 499.
(10) Pandya, P. R.; Singh, K. M.; Parnerkar, S.; Tripathi, A. K.; Mehta, H. H.; Rank, D. N.; Kothari, R. K.; Joshi, C. G. *J. Appl. Genet.* **2010**, *51* (3), 395.
(11) Dassa, B.; Borovok, I.; Lamed, R.; Henrissat, B.; Coutinho, P.; Hemme, C. L.; Huang, Y.; Zhou, J.; Bayer, E. a. *BMC Genomics* **2012**, *13* (1), 210.
(12) Hamberg, Y.; Ruimy-Israeli, V.; Dassa, B.; Barak, Y.; Lamed, R.; Cameron, K.; Fontes, C. M. G. a.; Bayer, E. a.; Fried, D. B. *PeerJ* **2014**, *2*, No. e636.
(13) Valbuena, A.; Oroz, J.; Hervás, R.; Vera, A. M.; Rodríguez, D.; Menéndez, M.; Sulkowska, J. I.; Cieplak, M.; Carrión-Vázquez, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13791.
(14) Mechaly, A.; Fierobe, H.-P.; Belaich, A.; Belaich, J.-P.; Lamed, R.; Shoham, Y.; Bayer, E. A. *J. Biol. Chem.* **2001**, *276* (13), 9883.
(15) Stahl, S. W.; Nash, M. a.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. a.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (50), 20431.
(16) Florin, E. L.; Moy, V. T.; Gaub, H. E. *Science* **1994**, *264* (5157), 415.
(17) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276* (5315), 1109.
(18) Rico, F.; Gonzalez, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. *Science* **2013**, *342* (6159), 741.
(19) King, G. M.; Carter, A. R.; Churnside, A. B.; Eberle, L. S.; Perkins, T. T. *Nano Lett.* **2009**, *9* (4), 1451.
(20) Müller, D. J.; Engel, A. *Nat. Protoc.* **2007**, *2* (9), 2191.
(21) Fernandez, J. M.; Li, H. *Science* **2004**, *303* (5664), 1674.
(22) Walder, R.; LeBlanc, M.-A.; Van Patten, W. J.; Edwards, D. T.; Greenberg, J. A.; Adhikari, A.; Okoniewski, S. R.; Sullan, R. M. A.; Rabuka, D.; Sousa, M. C.; Perkins, T. T. *J. Am. Chem. Soc.* **2017**, *139* (29), 9867.
(23) Otten, M.; Ott, W.; Jobst, M. a.; Milles, L. F.; Verdorfer, T.; Pippig, D. a.; Nash, M. a.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127.
(24) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. *Nano Lett.* **2015**, *15* (11), 7370.
(25) Seppälä, J.; Bernardi, R. C.; Haataja, T. J. K.; Hellman, M.; Pentikäinen, O. T.; Schulten, K.; Permi, P.; Ylänne, J.; Pentikäinen, U. *Sci. Rep.* **2017**, *7* (1), 4218.
(26) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. a.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. a.; Schulten, K.; Gaub, H. E.; Nash, M. a. *Nat. Commun.* **2014**, *5*, 5635.
(27) Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 81.
(28) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. *J. Phys. Chem. B* **2017**, *121*, 3620.
(29) Brand, U.; Gao, S.; Engl, W.; Sulzbach, T.; Stahl, S. W.; Milles, L. F.; Nesterov, V.; Li, Z. *Meas. Sci. Technol.* **2017**, *28* (3), 034010.
(30) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. *Nat. Protoc.* **2006**, *1* (1), 280.
(31) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72* (4), 1568.
(32) Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72* (4), 1541.
(33) Goh, B. C.; Hadden, J. A.; Bernardi, R. C.; Singharoy, A.; McGreevy, R.; Rudack, T.; Cassidy, C. K.; Schulten, K. *Annu. Rev. Biophys.* **2016**, *45*, 253.
(34) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Curr. Protoc. Protein Sci.* **2007**, 2.9.1.
(35) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215* (3), 403.
(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. In *Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; International Tables for Crystallography, Vol. *F*; Springer: Netherlands, 2006; pp 675–684.
(37) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781.
(38) Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. *Sci. Rep.* **2016**, *6*, 26536.
(39) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer, Berlin, Heidelberg, 1999; pp 39–65.
(40) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96* (10), 108101.

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841–17852

(41) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6620.
(42) Lange, O. F.; Grubmüller, H. *Proteins: Struct., Funct., Genet.* **2006**, *62* (4), 1053.
(43) Zhang, Y. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 342.
(44) Riederer, A.; Takasuka, T. E.; Makino, S.-I.; Stevenson, D. M.; Bukhman, Y. V.; Elsen, N. L.; Fox, B. G. *Appl. Environ. Microbiol.* **2011**, *77* (4), 1243.
(45) Han, S. O.; Yukawa, H.; Inui, M.; Doi, R. H. *J. Bacteriol.* **2003**, *185* (20), 6067.
(46) Chassard, C.; Delmas, E.; Robert, C.; Lawson, P. A.; Bernalier-Donadille, A. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 138.
(47) Moraïs, S.; Ben David, Y.; Bensoussan, L.; Duncan, S. H.; Koropatkin, N. M.; Martens, E. C.; Flint, H. J.; Bayer, E. A. *Environ. Microbiol.* **2016**, *18* (2), 542.
(48) Tsai, S.-L.; Goyal, G.; Chen, W. *Appl. Environ. Microbiol.* **2010**, *76* (22), 7514.

DOI: 10.1021/jacs.7b07574
J. Am. Chem. Soc. 2017, 139, 17841−17852

Supporting Information for

Combining *in Vitro* and *in Silico* Single Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

Tobias Verdorfer[1], Rafael C. Bernardi[2], Aylin Meinhold[1], Wolfgang Ott[1], Zaida Luthey-Schulten[2,3], Michael A. Nash*[4,5] and Hermann E. Gaub[1]

[1] Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany.
[2] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States.
[3] Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States.
[4] Department of Chemistry, University of Basel, 4056 Basel, Switzerland.
[5] Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland.

Author Contributions:
TV: experiment design, sample preparation, measurements, data analysis, writing of manuscript; RCB: SMD simulations and analysis, writing of manuscript; AM: sample preparation, measurements; WO: protein design and expression; ZLS: simulation analysis, writing of manuscript; MAN: experiment design, writing of manuscript; HEG: experiment design, writing of manuscript;

*Correspondence and requests for materials should be addressed to M.A.N. (email: michael.nash@unibas.ch).

KEYWORDS Single molecule force spectroscopy, cohesin-dockerin, cellulosome, steered molecular dynamics, network analysis

Materials and Methods

All reagents were at least of analytical purity grade and all buffers were filtered using a 0.2 µm polyethersulfone membrane filter (Nalgene, Rochester, NY, USA) prior to use. All incubation steps were done at room temperature, if not otherwise stated.

Gene construction, protein expression and purification

All genes were codon optimized for *E. coli* and synthesized (Invitrogen GeneArt Gene Synthesis - Thermo Fisher Scientific Messtechnik GmbH, Regensburg, Germany). All constructs were cloned into pET28a vectors using the Gibson assembly strategy[1] (New England Biolabs, MA, USA). All protein sequences can be found in the Supplementary Information.

ScaA cohesin mutant plasmid DNA was constructed using individually designed primers (Eurofins Genomics GmbH, Ebersberg, Germany) and the Phusion High-Fidelity PCR Kit (Thermo Fisher Scientific Messtechnik GmbH, Regensburg, Germany). The resulting double stranded linear DNA was ligated, phosphorylated and the template DNA was digested, in parallel, using a homemade reaction mix (1µl CutSMART buffer, 1µl ATP, 1µl T4 Polynucleotide Kinase, New England Biolabs, MA, USA, 1µl *Dpn*I, 1µl T4 DNA ligase and 0.5µl PEG-6k, Thermo Fisher Scientific GmbH, Regensburg, Germany, combined with 4.5µl unpurified PCR product) incubated at 37°C for 15min, 22°C for 45min and finally at 80°C for 5min. All plasmids used in *in vitro* protein expression were amplified in DH5-alpha cells, purified using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany), eluted with ultrapure water and stored at -20° C. All sequences were finally checked by DNA sequencing (Eurofins Genomics GmbH, Ebersberg, Germany).

Coh3-ddFLN$_4$-HIS-ybbR protein was expressed in *E. coli* NiCo21(DE3) cells (New England Biolabs, MA, USA). Precultures of 5 mL in LB medium, grown overnight at 37°C, were inoculated in ZYM-5052 auto-induction media containing kanamycin and grown for 6 h at 37°C followed by 24 h at 25°C[2]. Bacteria were spun down, and stored at -80°C. The pellet was resuspended and cells were lysed through sonication followed by centrifugation at 18000 g for 1 h at 4°C. The supernatant was applied to a Ni-NTA column (GE Healthcare, MA, USA) for HIS-Tag purification and washed extensively using HIS wash buffer (25mM TRIS, 500mM NaCl, 0.25% Tween-20, 10 % (v/v) Glycerol, 20mM imidazole, pH 8.5 @ 4°C), followed by a elution using HIS elution buffer (HIS wash buffer with 200mM imidazole instead of 20mM). Fractions containing protein were concentrated over regenerated cellulose filters (Amicon, Merck KGaA, Darmstadt, Germany), exchanged into measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl2, pH 7.2) using desalting columns (Zeba, Thermo Scientific, MA, USA), and frozen with 25 % (v/v) glycerol in liquid nitrogen to be stored at -80°C until used in experiments.

AFM Sample preparation

The sample preparation in these experiments follows in principle previously published protocols.[3–5] In brief, both the AFM cantilevers (Biolever Mini, Olympus, Tokyo, Japan) and the microscope slides (76mmx26mm, Carl Roth GmbH, Karlsruhe, Germany) were cleaned and silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, abcr GmbH, Karlsruhe, Germany) by baking at 80°C for 1h. A multiwell mask (CultureWell Gasket, Grace Bio-Labs, Bend, USA) was cleaned by sonication in a 1:1 mixture of isopropyl alcohol (IPA) and ultrapure water and then dried in a stream of nitrogen. The mask was attached to the glass slide to allow compartmentalization of the surface. The cantilevers were incubated with heterobifunctional NHS-PEG-Maleimide (5 kDa; Rapp Polymere, Tübingen, Germany) in 100mM HEPES buffer pH 7.4 for 45 min. The surfaces in the wells however were incubated with a 1:100

mixture of NHS-PEG-Maleimide and NHS-PEG-CH_3 (both 5 kDa; Rapp Polymere, Tübingen, Germany) in 100mM HEPES buffer pH 7.4, which, as experience has shown, will later result in the right surface density of immobilized protein for SMFS measurements in these experiments. After rinsing with ultrapure water, both the cantilevers and the surfaces were incubated with 1 mM Coenzyme A (CoA) in a 1 mM sodium phosphate pH 7.2, 50 mM NaCl, 10 mM EDTA buffer for at least 1 h. After a final ultrapure water rinse the cantilevers were incubated with 40µM Coh3-ddFLN$_4$-HIS-ybbR and 5 µM phosphopantetheinyl transferase (Sfp) for 2 h with magnesium chloride supplemented measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl2, 20mM $MgCl_2$ pH 7.2). The glass slide with the multiwell mask still attached was stored under Argon for later use. The cantilevers were rinsed extensively with measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl2, pH 7.2) and finally stored in it until use in measurement.

One-step *in vitro* expression and protein pulldown

PURExpress® IVTT-kit was thawed on ice and supplemented with 5 µM Sfp, 0.8 U/µl RNase inhibitor (NEB #M0314), 10 ng/µl Plasmid-DNA, 0.05% v/v Triton X-100 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) and ultrapure water resulting in total volumes of 10µl for each reaction mix. There was no need to supply this reaction mix with additional $MgCl_2$ for the Sfp coupling reaction, since the PURExpress® IVTT-kit already contains 13mM $MgCl_2$[6]. The reaction mixes were transferred to the wells onto the CoA functionalized glass slide and incubated at 37°C for 3h. During this time the cell free expression kit is constantly producing proteins, while at the same time the Sfp couples expressed protein to the surface via the ybbR tags. Both cantilevers and surfaces were rinsed extensively with measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl2, pH 7.2) before measurement and finally the multiwell mask was removed from the surface and stored in a 1:1 mixture of IPA and ultrapure water for further use.

AFM SMFS measurements

A custom build AFM connected to a MFP3D controller (Asylum Research, Santa Barbara, CA, USA) was used for all measurements. Acquisition- and instrument control software was written in Igor Pro 6 (Wavemetrics, OR, USA). The cantilever was aligned to each measurement spot by moving the sample using a 25mmx25mm piezomotor stage (PI, Karlsruhe, Germany) using a camera mounted below the sample. These positions were saved in the software for later use. The cantilever was brought in close proximity to the surface and constant speed measurements with retraction speeds of 1600 nm/s were started. The glass surface was moved horizontally by 100 nm in a snail-like-pattern within each protein spot. After 2000 approach- and retract-cycles the AFM-head was automatically lifted by a linear piezo actuator (Newport, CA, USA) and the surface was moved horizontally by typically ~300 µm to expose the cantilever to the next protein spot. Cantilevers were calibrated using the equipartition theorem method[7].

AFM SMFS Data Analysis

Data analysis was carried out following previous work[8]. In short, data were transformed into physical units and corrected for cantilever bending, laser spot- and baseline-drift. Force peaks and rupture events were detected and transformed to contour length space. The Worm Like Chain model (WLC)[9] was used to fit relevant peaks. All curves showing a ddFLN4 and cohesin contour length increment (ddFLN4: 34nm[8,10], cohesin: 45nm) were used to assemble unfolding force histograms, which were then fitted following the Bell-Evans model[11,12], which is commonly used to estimate the distance to the transition state Δx and the natural off-rate k_0 of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments.

Bell-Evans probability density function at given loading rate r:

$$p(F) = \frac{k_0}{r}\, exp[\frac{\Delta x}{k_B T}F - \frac{k_0\, k_B T}{r\, \Delta x}(e^{\frac{\Delta x}{k_B T}F} - 1)]$$

The Bell-Evans model predicts a linear dependence between the most probable rupture force <F> and the logarithm of the force loading rate r:

$$< F(r) >= (\frac{k_B T}{\Delta x})\, ln(\frac{r\, \Delta x}{k_0\, k_B T})$$

The Dudko-Hummer-Szabo model[13] describes a non-linear dependence for the most probable rupture force on loading rate:

$$< F(r) >= \frac{\Delta G}{\nu \Delta x}\left\{1 - \left[\frac{k_B T}{\Delta G} ln\left(\frac{k_0\, k_B T}{r\, \Delta x} exp\left(\frac{\Delta G}{k_B T} + \gamma\right)\right)\right]^{\nu}\right\}$$

where ΔG is the free energy of activation and $\gamma \approx 0.577...$ is the Euler-Mascheroni constant. The model parameter ν defines the single-well free-energy surface model used (ν = 2/3 for linear-cubic and ν = 1/2 for cusp free-energy).

Structural Model Determination

The amino acid sequence of all seven cohesins under investigation were obtained from the GenBank (GenBank: AAF06064.1) proteomic server[14,15]. The template search was performed employing the similarity search algorithm in the protein Blast server (http://blast.ncbi.nlm.nih.gov/Blast.cgi)[16] using the Protein Data Bank[17] (http://www.pdb.org) as database and the default options. Using VMD's[18] multiseq[19] analysis tool, sequences were aligned to templates employing ClustalW algorithim[20]. The construction of cohesin models were performed using MODELLER 9.17 software[21] that employs spatial restriction techniques based on the 3D-template structure. The best model was selected by analyzing the stereochemical quality check using PROCHECK[22] and overall quality by ERRAT server.[23] All structures were subjected to 100 ns of equilibrium MD, as described below, to ensure conformational stability. All structures shown in this manuscript are from post-equilibration simulations.

Molecular dynamics simulations

Employing advanced run options of QwikMD,[24] structural models were solvated and the net charge of the proteins were neutralized using a 75 mM salt concentration of sodium chloride, which were randomly arranged in the solvent. The overall number of atoms included in MD simulations varied from 50,000 in the equilibrium simulations to near 300,000 in the pulling simulations. The MD simulations in the present study were performed employing the NAMD molecular dynamics package.[25] The CHARMM36 force field[26,27] along with the TIP3 water model[28] was used to describe all systems. The simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME)[29] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[25] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations all the systems were submitted to an energy minimization protocol for 1,000 steps. MD simulations with position restraints in the protein backbone atoms were performed for 10 ns and served to pre-equilibrate systems before the 100 ns equilibrium MD runs, which served to evaluate structural model stability. During the 10 ns pre-equilibration the initial temperature was set to zero and was constantly increased by 1 K every 1,000 MD steps until the desired temperature (300 K) was reached.

With structures properly equilibrated and checked, solvent boxes were enlarged in the Z coordinate to allow space for protein unfolding during SMD simulations. The new solvent boxes were equilibrated for 10 ns keeping the protein atoms restrained in space. SMD simulations[11] were performed using a constant velocity stretching (SMD-CV protocol), employing four different pulling speeds: 250, 25, 2.5 and 0.5 Å/ns.

Simulation replicas (at least 25 per system), used in all the plots in this manuscript, were performed with constant pulling speed of 2.5 Å/ns. Values for force over the pulling spring were saved every 50 steps. The spring constant of the pulling spring was set to 5.0 kcal/mol/Å^2, while the holding spring had a constant of 10 kcal/mol/Å^2. In all simulations, totaling over 350 SMD simulations, SMD was employed by harmonically restraining the position of N-terminal amino acid residue of the cohesin domain, and moving a second restraint point, at the C-terminal of the cohesin domain, with constant velocity in the +z direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Owing to the flexibility of the linkers between the cohesins and fingerprint domains, this approach reproduces the experimental set-up. All analyses of MD trajectories were carried out employing VMD[18] and its plug-ins. Secondary structures were assigned using the Timeline plug-in, which employs STRIDE criteria.[30]

The Network View plugin[31] on VMD was employed to perform dynamical network analysis. A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 non-consecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (non-hydrogen) from the 2 monomers are within 4.5Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.[31] nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths, which can be understood as allosteric signaling pathways or force propagation pathways [32]. Suboptimal paths are defined as paths that are slightly longer than the optimal path, with a given suboptimal path visiting a node not more than once. These multiple communication paths are nearly equal in length, and not all residues along these paths need be considered important for allostery. Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective pathway for allosteric communication. The thickness of the edges connecting the nodes reveals the least and most used paths. Allostery can be understood in terms of pathways of residues that efficiently transmit energy, here in the form of mechanical stress, between different binding sites[33]. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as wij = −log (| Cij |), where Cij is the correlation matrix calculated with Carma[34]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was −log (0.5) = 0.69. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently. As previously investigated by our group [32], Pearson correlation is ideal for force propagation calculation. However, due to its nature, communities analysis would benefit from an information-theory-based method, so here we employed generalized correlation[35] to the community analysis. Tightly correlated groups of atoms are clustered into communities, indicating functional domains of biomolecules and important interfaces between multi-molecule complexes.

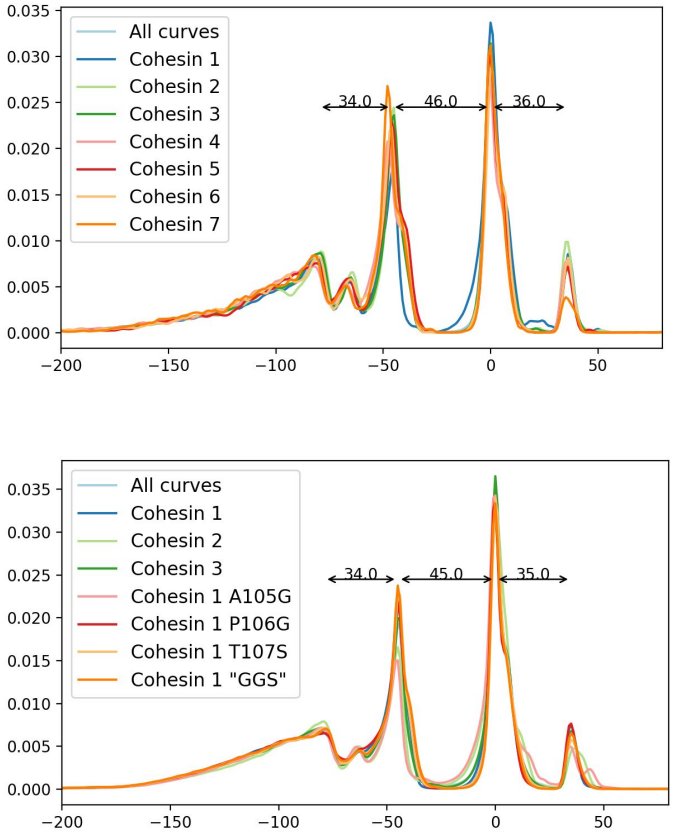


Figure S1: Frequency of observed relative contour lengths increments determined by transforming multiple force traces into contour length space via the worm-like chain model and aligning them. The individual increments (f.l.t.r.) correspond to the unfolding of the ddFLN4 fingerprint domain, the ScaA cohesins and the occasional unfolding of the X-module of the Coh3.XDoc3 complex[36].

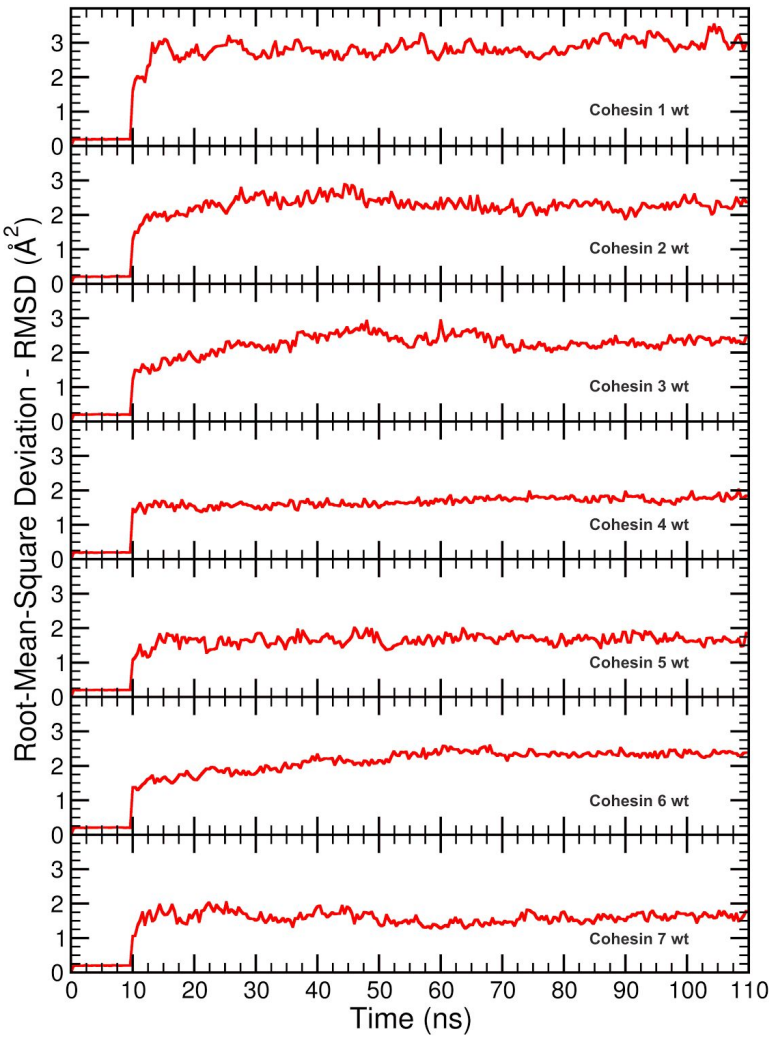


Figure S2: Root-Mean-Square Deviation (RMSD) for equilibrium simulations. All constructs were simulated with position restraints of the backbone atoms during 10 ns and free of restraints during 100 ns. All plots show stable structures after approximately 30ns. It is noteworthy that hanging cohesins have a higher RSMD value, particularly cohesin 1.

S7

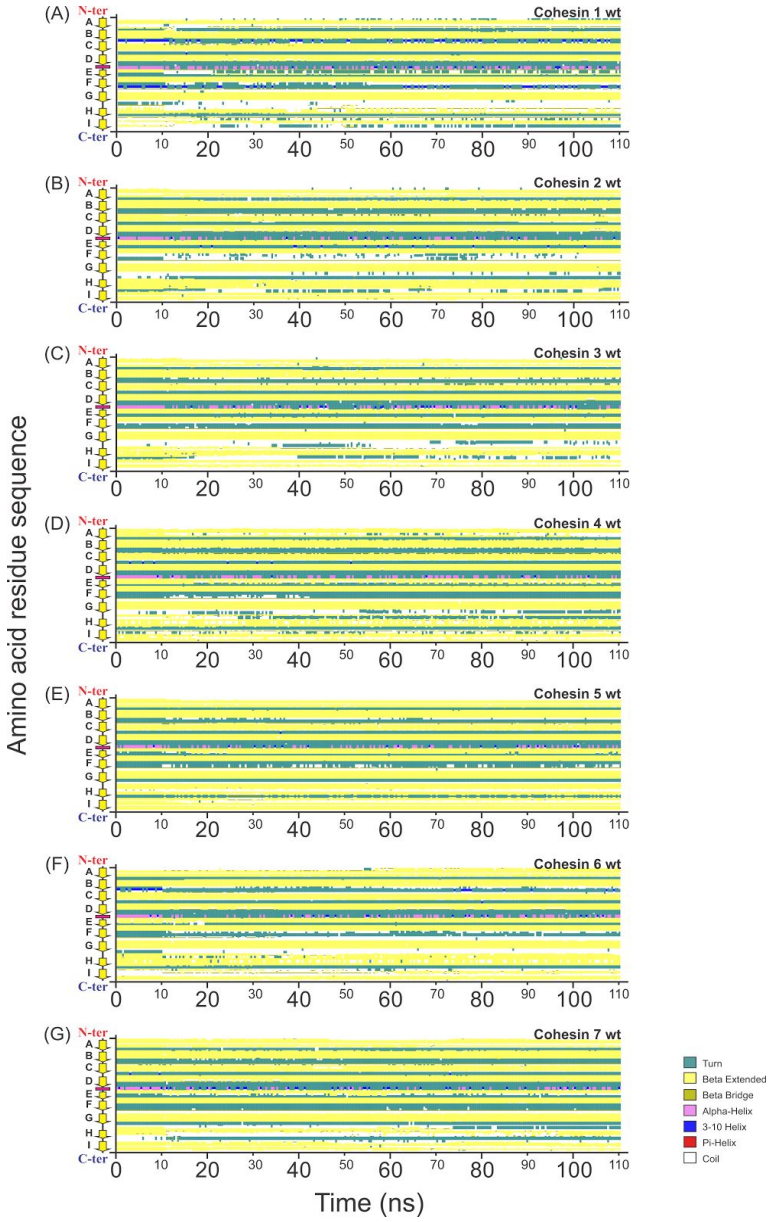


Figure S3: Secondary Structure evolution during equilibration simulations. Secondary structure content was evaluated using VMD's Timeline during the equilibration simulations. All constructs were simulated with position restraints of the backbone atoms during 10 ns and free of restraints during 100 ns. All conformations show stable structures.

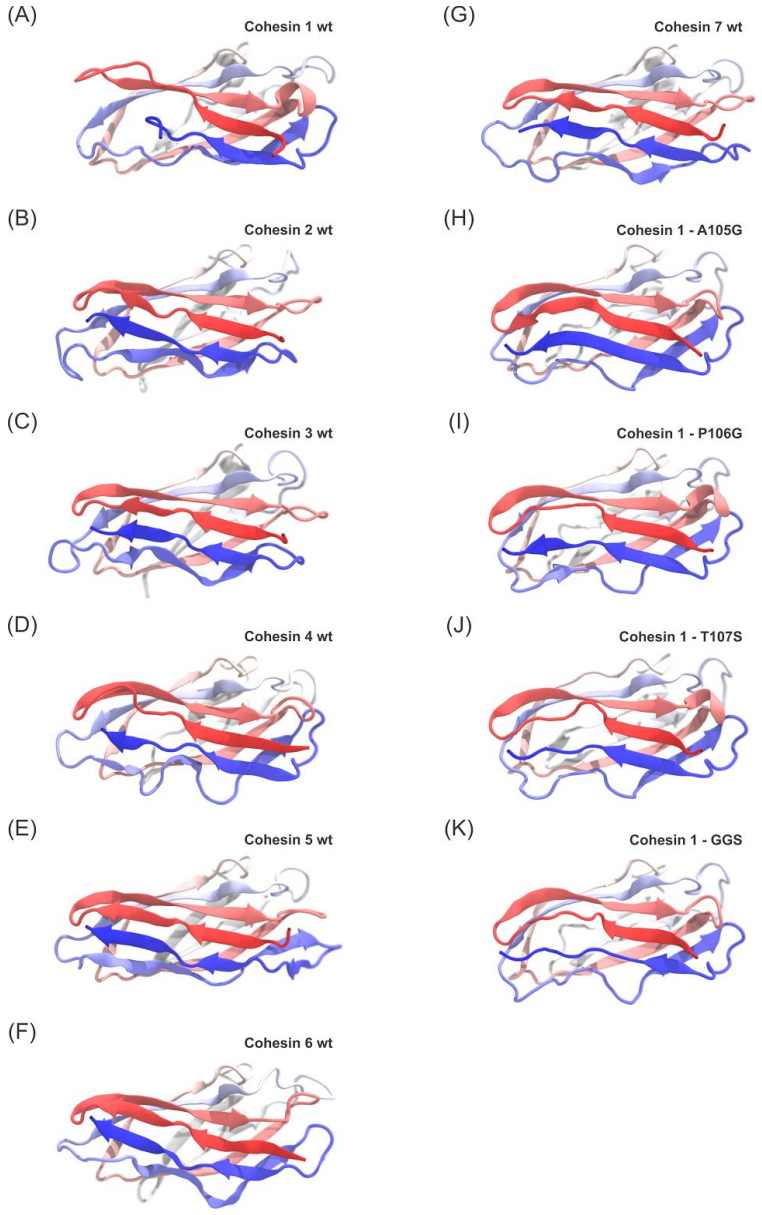


Figure S4. Structural model of studied cohesins after 100ns of MD simulation. All structures were obtained using Modeller 9.17 and subjected to 100 ns of molecular dynamics equilibration using QwikMD and NAMD. All images were prepared using VMD.

S9

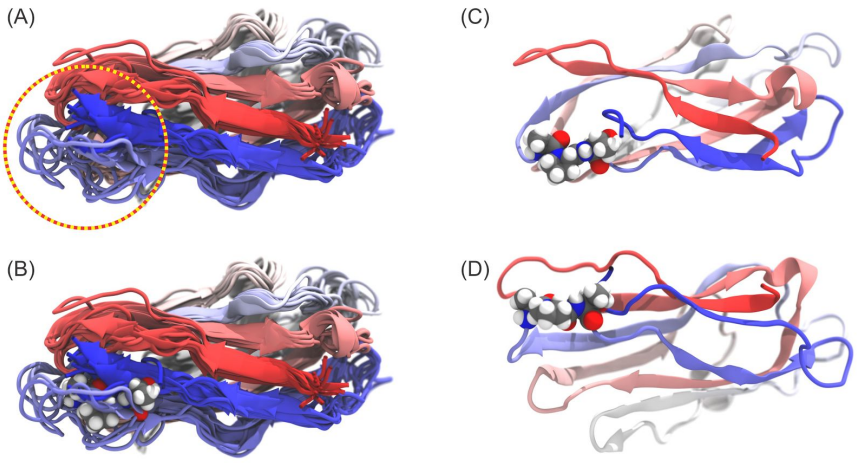


Figure S5. Structural model for ScaA cohesins after 100ns of MD simulation. (A) Using Modeller, very similar model structures were obtained for ScaA cohesins. The region highlighted in the circle was observed to be the most flexible one in cohesin 1, presenting a different behavior than the other cohesins. (B) Sequence alignment shows that, in the highlighted region of (A), 3 amino acid residues of cohesin 1 were different, compared to cohesin 2 and 3, namely ALA105, PRO106 and THR107. (C) (D) Two different viewpoints of cohesin 1 with highlighted ALA105, PRO106 and THR107.

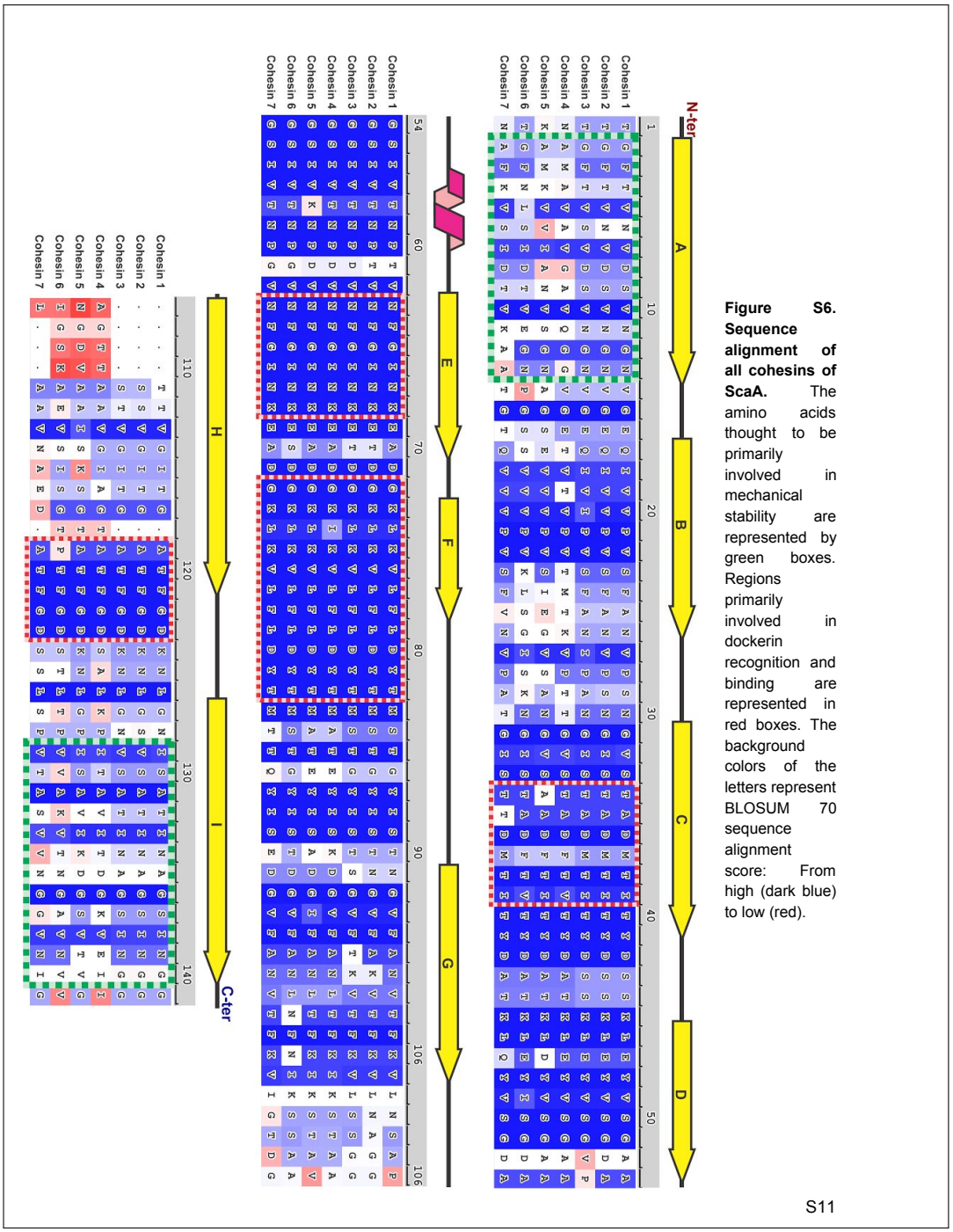


Figure S6. Sequence alignment of all cohesins of ScaA. The amino acids thought to be primarily involved in mechanical stability are represented by green boxes. Regions primarily involved in dockerin recognition and binding are represented in red boxes. The background colors of the letters represent BLOSUM 70 sequence alignment score: From high (dark blue) to low (red).

S11

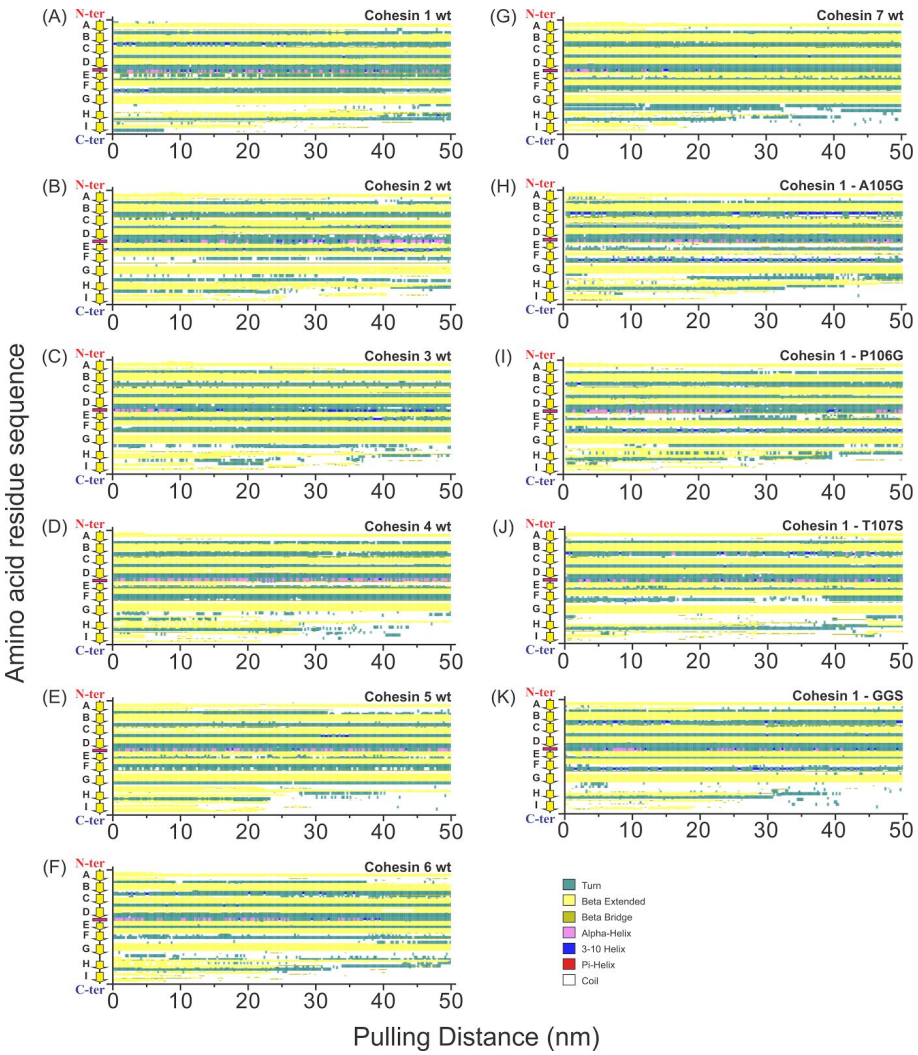


Figure S7. Secondary Structure evolution during pulling simulations. Secondary structure content was evaluated using VMD's Timeline during the SMD simulations. Here we show the evolution of the secondary structure during the first 50nm of pulling, which corresponds to the region where the peak force is observed. The plots present a representative simulation (one of the replicas) for each system studied. In all simulations presented the C-terminal region is the first to unfold, showing that the highest peak corresponds to losing the last beta-strand structure.

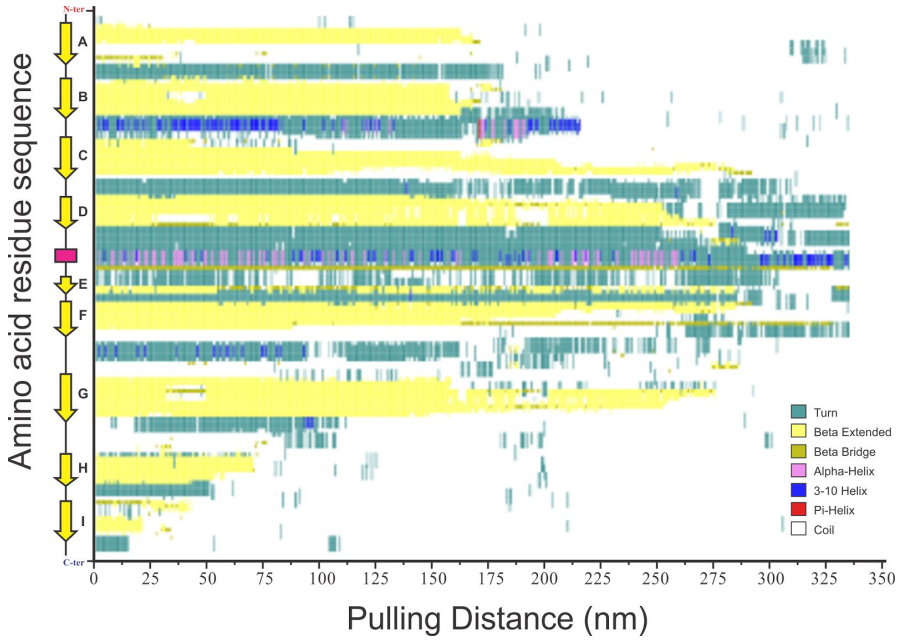


Figure S8. Cohesin 1 secondary structure evolution during pulling simulations. Secondary structure content was evaluated using VMD's Timeline during the SMD simulations. Here we show the evolution of the secondary structure during the whole unfolding process.

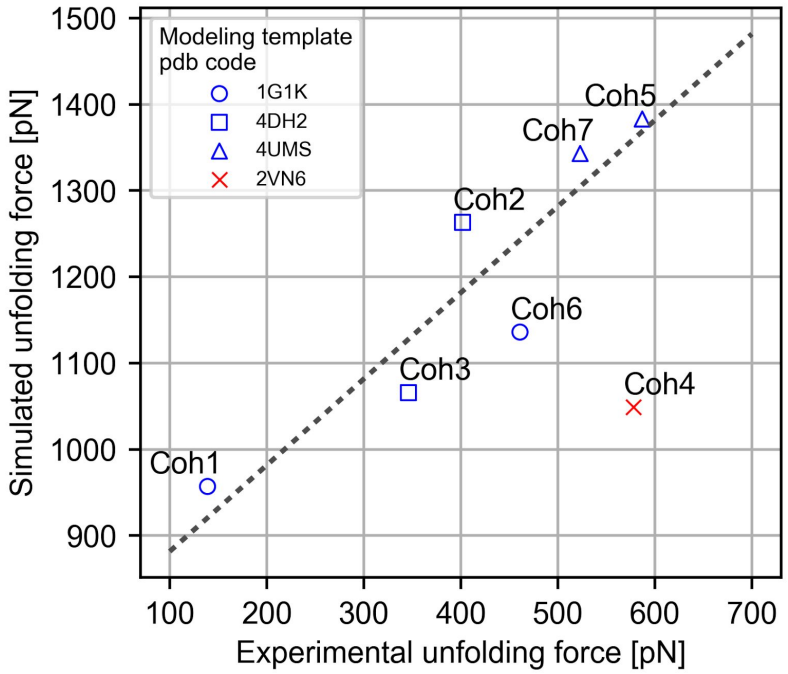


Figure S9. Experimental unfolding force vs. simulated peak unfolding force of all wild type cohesins. A linear fit shows a clear trend between measured and simulated unfolding forces with offset fitting parameter (±SD) F_0 = 782 ± 29 pN. Cohesin 4 (red) was excluded from the fit since, as explained in the main text, its homology modeling template was a suboptimal and likely resulted in a non-ideal initial folded state.

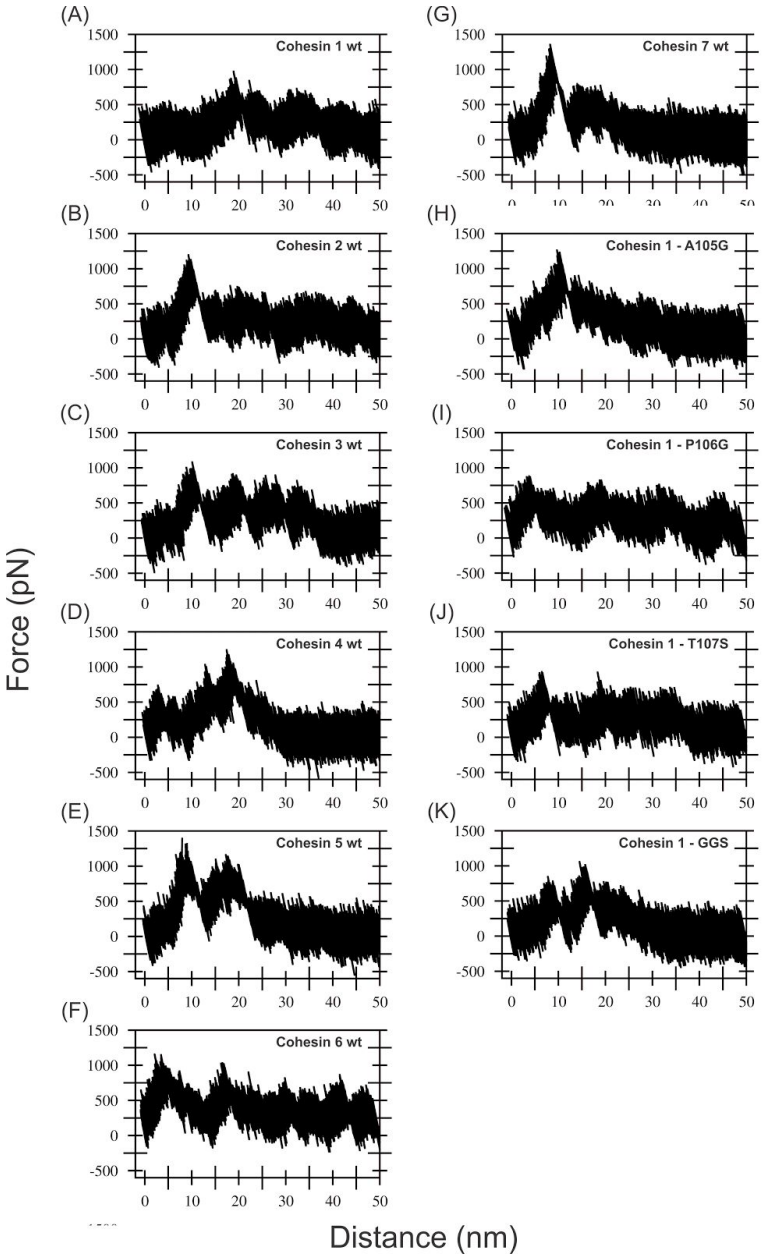


Figure S10. Force profile during pulling simulations. For the first 50 nm of pulling, the plots present a representative simulation (one of the replicas) for each system studied.

	Coh1	Coh2	Coh3	Coh4	Coh5	Coh6	Coh7
Coh1	1	**** 4.7E-11	** 0.00949	** 0.0072	**** 1.13E-10	**** 1.61E-07	**** 1.13E-10
Coh2	**** 4.7E-11	1	**** 4.14E-08	**** 1.51E-07	** 0.004	**** 1.28E-05	0.0717
Coh3	** 0.00949	**** 4.14E-08	1	0.985	**** 3.5E-09	** 0.00222	**** 1.72E-08
Coh4	** 0.0072	**** 1.51E-07	0.985	1	**** 2.04E-10	** 0.00233	**** 9.36E-09
Coh5	**** 1.13E-10	** 0.004	**** 3.5E-09	**** 2.04E-10	1	**** 8.55E-11	0.124
Coh6	**** 1.61E-07	**** 1.28E-05	** 0.00222	** 0.00233	**** 8.55E-11	1	**** 4.15E-07
Coh7	**** 1.13E-10	0.0717	**** 1.72E-08	**** 9.36E-09	0.124	**** 4.15E-07	1

*: p < 0.05 **: p < 0.01 ***: p < 0.001 ****: p < 0.0001

Figure S11. Statistical significance between the simulated unfolding forces of all wild type cohesins. P-values were calculated using the Kolmogorov-Smirnov test.

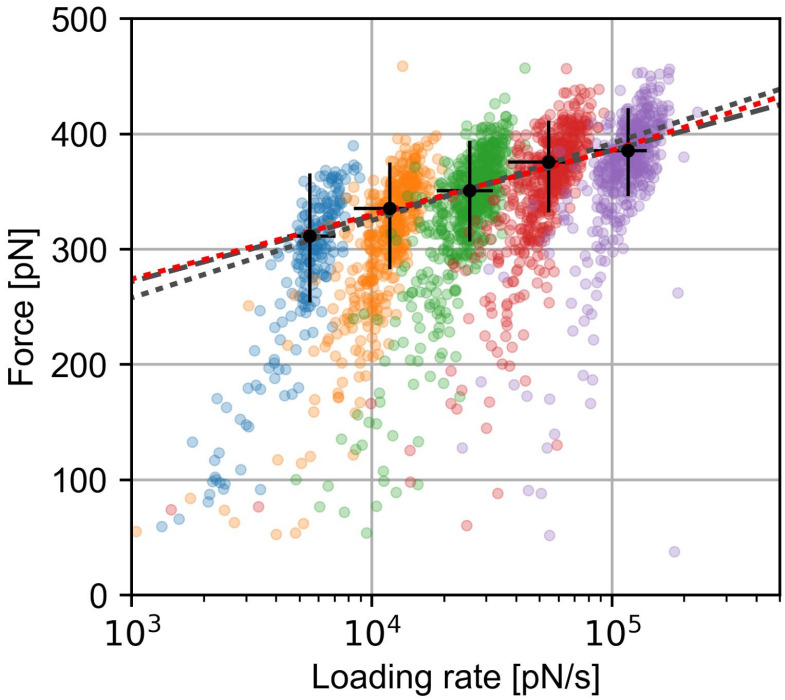


Figure S12. Experimental dynamic force spectrum for unfolding events of cohesin 3. Varicolored points represent rupture force/loading rate data from an experiment with 5 different pulling speeds. Black points represent the most probable rupture force/loading rate of each pulling speed obtained from kernel density estimates. Error bars represent the full width at half maximum. Gray lines represent least-squares fits of the Bell-Evans model to the experimental, and to both the experimental and the simulation data with fitting parameters (±SD) Δx=0.17 ± 0.012 nm, k_0=6.7 · 10^{-4} ± 6.3 · 10^{-4} s^{-1} and Δx=0.14 ± 0.0015 nm, k_0=4.9 · 10^{-3} ± 8.9 · 10^{-4} s^{-1}, respectively. The red dotted line represents a least-squares fit of the DHS model to both the combined experimental and the simulation data with fitting parameters (±SD) of Δx=0.19 ± 0.024 nm, k_0=1.4 · 10^{-4} ± 2 · 10^{-4} s^{-1} and ΔG=60 ± 13 k_BT. In this range of loading rates the Bell-Evans fit through the experimental data falls along the DHS fit.

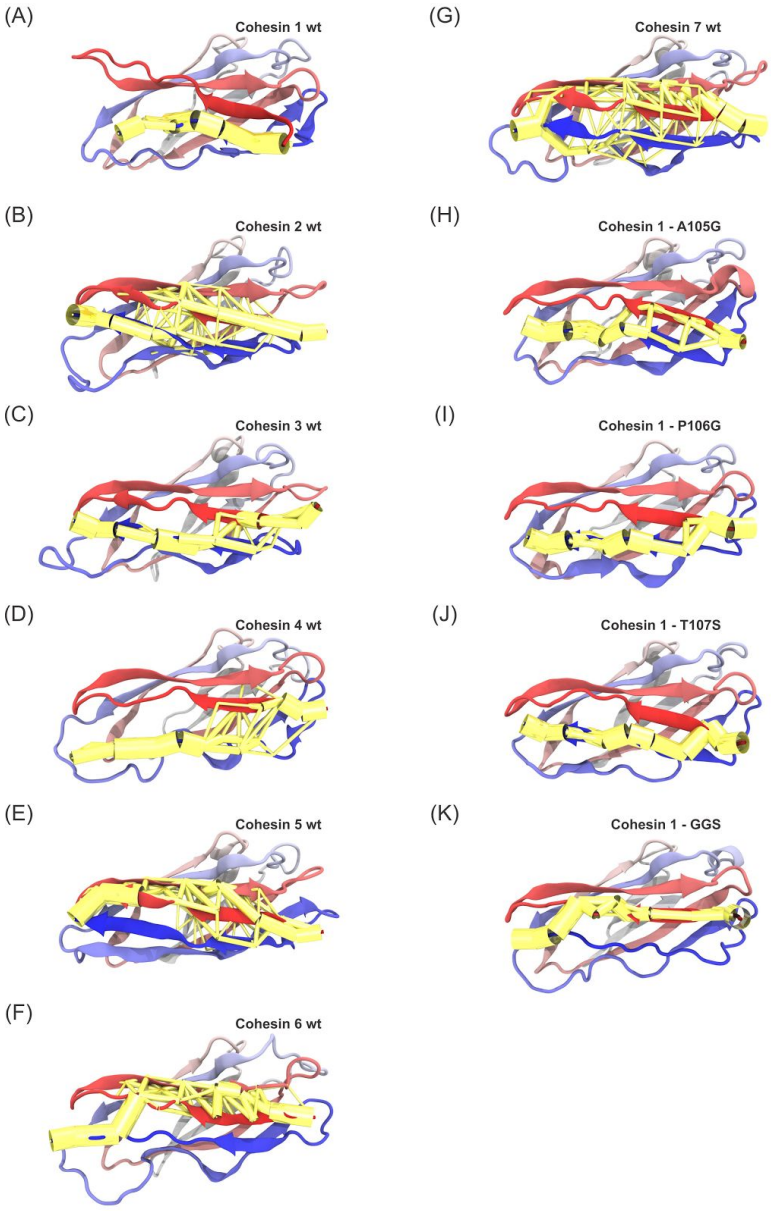


Figure S13. Network-based force propagation analysis calculated using Pearson correlation. Suboptimal force paths were calculated using VMD during the first 5nm of pulling simulation. Note that for all the systems beta-strands A,B and I are the main regions involved in the force propagation.

S18

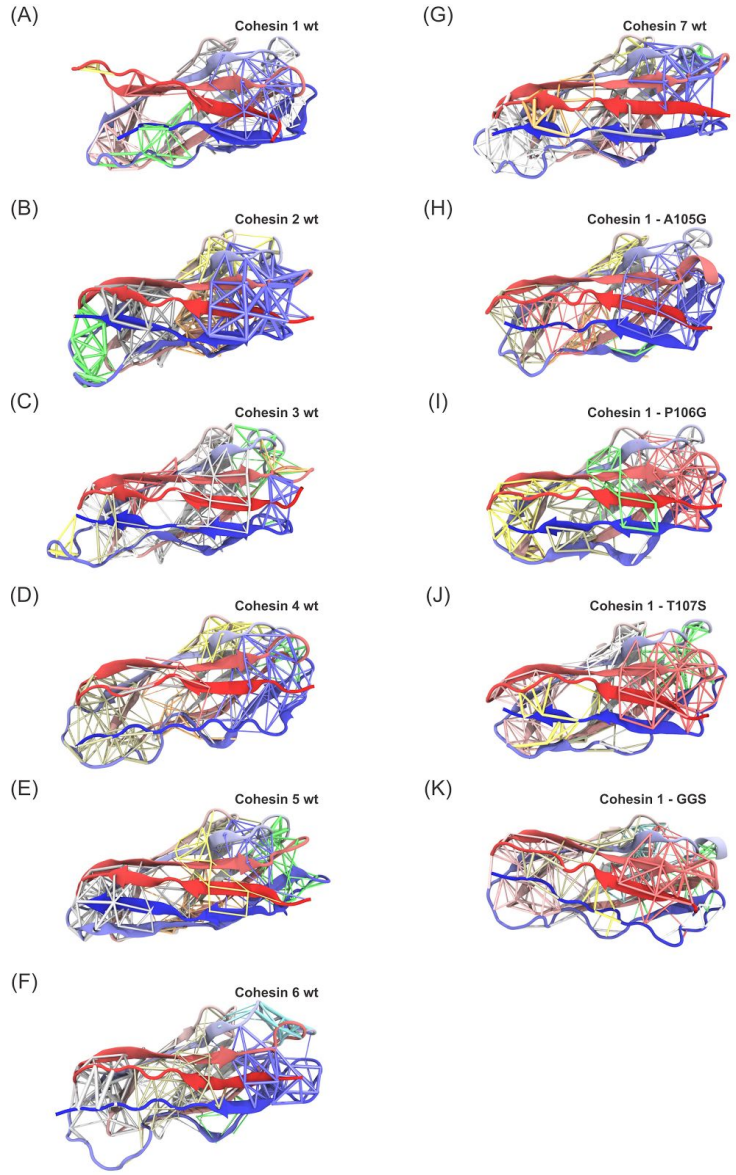


Figure S14. Network-based community analysis calculated using generalized correlation. Communities were calculated using VMD during the first 5nm of pulling simulation. Each color represents a different community. Colors of the communities in different systems are not related, and should not be compared as being the same community in different systems. Thickness of the network represents the *log* of the normalized correlation value. Thick connections represent highly correlated regions.

S19

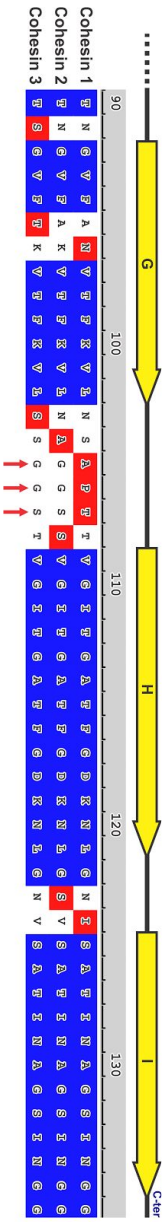


Figure S15. High resolution version of the sequence alignment of the C-terminal end of the hanging cohesins 1-3.

	Coh1	A105G	P106G	T107S	GGS
Coh1	1	* 0.0333	0.523	0.714	*** 0.000295
A105G	* 0.0333	1	** 0.00128	0.124	** 0.00999
P106G	0.523	** 0.00128	1	0.414	**** 1.33E-05
T107S	0.714	0.124	0.414	1	*** 0.000235
GGS	*** 0.000295	** 0.00999	**** 1.33E-05	*** 0.000235	1

*: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$ ****: $p < 0.0001$

Figure S16. Statistical significance between the simulated unfolding forces Cohesin 1 and its four mutants. P-values were calculated using the Kolmogorov-Smirnov test.

Protein Sequences

ybbR-tag - linker and additional residues - Cohesin - linker - XDoc3

MGTDSLEFIASKLALEVLFQGPLQHHHHHHPWTSAS

Cohesin 1
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFGINKEA
DGKLKVLFLDYTMSTGYISTNGVFANVTFKVLNSAPTTVGITGATFGDKNLGNISATINAGSINGG

Cohesin 2
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGISTADMTITYDSSKLEYVSGDAGSIVTNPTVNFGINKETD
GKLKVLFLDYTMSTGYISTNGVFAKVTFKVLNAGGSSVGITGATFGDKNLGSVSATINAGSINGG

Cohesin 3
TGFTVSVDSVNGNVGEQIVIPVSFANIPANGISTADMTITYDSSKLEYVSGVPGSIVTNPDVNFGINKETDG
KLKVLFLDYTMSTGYISTSGVFTKVTFKVLSSGGSTVGITGATFGDKNLGNVSATINAGSINGG

Cohesin 4
NAMAVAVGAVQGGVGETVTVPVTMTKVPTTGVSTADFTVTYDATKLEYVSGAAGSIVTNPDVNFGINKEA
DGKIKVLFLDYTMATEYISKDGVFANLTFKIKSTAAAGTTAAVGIAGTATFGDSALKPITAVITDGKVEII

Cohesin 5
KAMKVVIANVSGNAGSEVVVPVSIEGVSANGVSAADFTITYDATKLDYVSGAAGSIVKNPDVNFGINKEAD
GKLKVLFLDYTMATEYISADGIFANLTFKIKSTAVNGDVAAISKSGTATFGDKNLGPISAVIKDGSVTVG

Cohesin 6
TGFNLSIDTVEGNPGSSVVVPVKLSGISKNGISTADFTVTYDATKLEYISGDAGSIVTNPGVNFGINKESDG
KLKVLFLDYTMSTGYISTDGVFANLNFNIKSSAAIGSKAEVSISGTPTFGDSTLTPVVAKVTNGAVNVV

Cohesin 7
NAFKVSIDTVKAATGTQVVVPVSFVNVPATGISTTDMTITYDATKLQYVSGDAGSIVTNPGVNFGINKEAD
GKLKVLFLDYTMTTQYISEDGVFANVTFKVIGTDGLAAVNAEDATFGDSSLSPVTASVVNGGVNIG

Cohesin 1 A105G
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFGINKEA
DGKLKVLFLDYTMSTGYISTNGVFANVTFKVLNS**G**PTTVGITGATFGDKNLGNISATINAGSINGG

Cohesin 1 P106G
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFGINKEA
DGKLKVLFLDYTMSTGYISTNGVFANVTFKVLNSA**G**TTVGITGATFGDKNLGNISATINAGSINGG

Cohesin 1 T107S
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFGINKEA
DGKLKVLFLDYTMSTGYISTNGVFANVTFKVLNSAP**S**TVGITGATFGDKNLGNISATINAGSINGG

Cohesin 1 A105G P106G T107S
TGFTVNVDSVNGNVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFGINKEA
DGKLKVLFLDYTMSTGYISTNGVFANVTFKVLNS**GGS**TVGITGATFGDKNLGNISATINAGSINGG

VVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVYNEGYTGDDGVAVVLREYESEPVDITA
ELTFGDATPANTYKAVENKFDYEIPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYY
AATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYT
KSSVDQYKDMAANEPNKLWDIVTGDAEEE

S22

Coh3 - linker - ddFLN4 - linker and additional residues - ybbR

MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGYHIYWD
ERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTVELKVPADAKAGDVY
PIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGS
VVPSTGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNG
DGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSELKLPRSRHHHHHHGSLEVLFQGPDSL
EFIASKLA

References

(1) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. a.; Smith, H. O.; Iii, C. A. H.; America, N. *Nat. Methods* **2009**, *6* (5), 343.
(2) Studier, F. W. *Protein Expr. Purif.* **2005**, *41* (1), 207.
(3) Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. *J. Vis. Exp.* **2013**, No. 82, 1.
(4) Otten, M.; Ott, W.; Jobst, M. a.; Milles, L. F.; Verdorfer, T.; Pippig, D. a.; Nash, M. a.; Gaub, H. E. *Nat. Methods* **2014**, No. september, 1.
(5) Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K. *Nat. Protoc.* **2010**, *5* (6), 975.
(6) Shimizu, Y.; Kanamori, T.; Ueda, T. *Methods* **2005**, *36* (3), 299.
(7) Hutter, J. L.; Bechhoefer, J. *Rev. Sci. Instrum.* **1993**, *64* (7), 1868.
(8) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. *J. Phys. Chem. B* **2016**, acs.jpcb.6b09593.
(9) Siggia, E. D.; Bustamante, C.; Marko, J. F.; Smith, S. B. *Science* **1994**, 5.
(10) Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 81.
(11) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72* (4), 1568.
(12) Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72* (4), 1541.
(13) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96* (10), 108101.
(14) Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. *Nucleic Acids Res.* **2017**, *45* (D1), D37.
(15) Ding, S.-Y.; Bayer, E. A.; Steiner, D.; Shoham, Y.; Lamed, R. *J. Bacteriol.* **1999**, *181* (21), 6720.
(16) Gish, W.; States, D. J. *Nat. Genet.* **1993**, *3* (3), 266.
(17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; International Tables for Crystallography; Springer Netherlands, 2006; pp 675–684.
(18) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33.
(19) Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. *BMC Bioinformatics* **2006**, *7*, 382.
(20) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Curr. Protoc. Bioinformatics* **2002**, *Chapter 2*, Unit 2.3.
(21) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Curr. Protoc. Protein Sci.* **2007**, *Chapter 2*, Unit 2.9.
(22) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26* (2), 283.
(23) MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. *Curr. Opin. Struct. Biol.* **1994**, *4* (5), 731.
(24) Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. *Sci. Rep.* **2016**, *6*, 26536.

(25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781.
(26) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257.
(27) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102* (18), 3586.
(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926.
(29) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089.
(30) Frishman, D.; Argos, P. *Proteins* **1995**, *23* (4), 566.
(31) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6620.
(32) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. *Nano Lett.* **2015**, *15* (11), 7370.
(33) Ribeiro, A. A. S. T.; Ortiz, V. *Chem. Rev.* **2016**, *116* (11), 6488.
(34) Glykos, N. M. *J. Comput. Chem.* **2006**, *27* (14), 1765.
(35) Lange, O. F.; Grubmüller, H. *Proteins* **2006**, *62* (4), 1053.
(36) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. a.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. a.; Schulten, K.; Gaub, H. E.; Nash, M. a. *Nat. Commun.* **2014**, *5*, 5635.

Chapter 5

Multiplexing of AFM Experiments

5.1 Summary

Cell-free molecular biology allows high throughput protein production on a smaller scale [131]. Since single molecule techniques operate in a nanomolar concentration range cell-free approaches are a good match. With robust and efficient immobilization techniques it is possible to localize proteins with high densities at specific points, allowing the use of microfluidics to increase throughput in atomic force microscope (AFM) experiments. **Associated publication P6** describes an approach for multiplexed AFM studies.
A PDMS (Polydimethylsiloxane)-based microfluidics chip contains several hundred spatially separated chambers, where proteins can be synthesized, immobilized and analyzed. The microfluidics chip is designed such that an experimental compartment consists of two units [132]. A back-chamber where the DNA is localized and a front-chamber where protein immobilization takes place. First the glass slide is functionalized and passivated for site-directed pull-down capabilities. Then a cell-free extract is flushed into the back-chamber to transcribe DNA to mRNA and to translate it into a protein. The DNA encodes immobilization-tag, protein of interest and binding domain. Immobilization-tag and binding domain are at opposite termini to ensure only fully translated proteins are probed later on one surface with the same cantilever.
After protein synthesis, the spatially separated chambers (back and front) are connected by opening the separating valve and the newly synthesized proteins diffuse to the activated surface in the front. The microfluidics chip that served as reaction compartment can now be removed and the glass slide can be mounted under an AFM head, presenting an array of individually and spatially separated proteins of interest to be probed on one surface.
In this study different fingerprint domains were fused with a type I dockerin (from *Clostridium thermocellum*) and probed with a cantilever having a type I cohesin attached. Several proteins could be probed with the same cantilever over a longer period of time, increasing comparability of absolute unfolding forces of the different fingerprint domains.

5.2 Associated Publication P6

From genes to protein mechanics on a chip

by

Marcus Otten*, Wolfgang Ott*, Markus A. Jobst*, Lukas F. Milles,

Tobias Verdorfer, Diana A. Pippig, Michael A. Nash and Hermann E. Gaub

*(*contributed equally)*

published in

Nature Methods, 11 (11), 1127-1130, (2014)

Reprinted from [133], with permission from Nature Publishing Group.

BRIEF COMMUNICATIONS

From genes to protein mechanics on a chip

Marcus Otten[1,2,4], Wolfgang Ott[1,2,4], Markus A Jobst[1,2,4], Lukas F Milles[1,2], Tobias Verdorfer[1,2], Diana A Pippig[1–3], Michael A Nash[1,2] & Hermann E Gaub[1,2]

Single-molecule force spectroscopy enables mechanical testing of individual proteins, but low experimental throughput limits the ability to screen constructs in parallel. We describe a microfluidic platform for on-chip expression, covalent surface attachment and measurement of single-molecule protein mechanical properties. A dockerin tag on each protein molecule allowed us to perform thousands of pulling cycles using a single cohesin-modified cantilever. The ability to synthesize and mechanically probe protein libraries enables high-throughput mechanical phenotyping.

Mechanical forces play a pivotal role in biological systems by performing tasks such as guiding cell adhesion[1], inducing gene expression patterns[2] and directing stem cell differentiation[3]. At the molecular level, mechanosensitive proteins act as sensors and transducers, communicating the presence and direction of applied forces to downstream signaling cascades. Conformational changes in response to mechanical forces[4] and energetic barriers along unfolding pathways can be probed by single-molecule force spectroscopy (SMFS) techniques[4]. Such techniques, including optical tweezers, magnetic tweezers and atomic force microscopy (AFM), have been used to interrogate high-affinity receptor-ligand binding[5], measure unfolding and refolding dynamics of individual protein domains[6–8], observe base-pair stepping of RNA polymerases[9] and identify DNA stretching and twisting moduli[10].

Despite these successes, SMFS experiments have been limited by low throughput. Experimental data sets typically contain a majority of unusable force-distance traces owing to the measurement of multiple molecular interactions in parallel or a lack of specific interactions. Typical yields of interpretable single-molecule interaction traces in SMFS experiments vary between 1% and 25%. The incapacity of SMFS to quickly screen libraries of molecular variants has hindered progress toward understanding sequence-structure-function relationships at the single-molecule level. In particular, the need to prepare each protein sample and cantilever separately increases experimental workload and gives rise to calibration uncertainties. Therefore, methods to interrogate the mechanical behavior of different proteins in a parallel and streamlined format with the same cantilever would offer distinct advantages. Such a screening approach could characterize single-molecule properties such as unfolding forces, interdomain mechanical signatures and mechanically activated catch-bond behavior[1]. Screening of these properties could find applications in biotechnology and human health studies in which mechanical dysregulation or misfolding is suspected to play a role in pathology[11].

Here we developed a platform for parallel characterization of individual protein mechanics in a single experiment (**Fig. 1**). Microspotted gene arrays were used to synthesize fusion proteins *in situ* using cell-free gene expression. Proteins were covalently immobilized inside multilayer microfluidic circuits. A single cantilever was then positioned above the protein array and used to probe the mechanical response of each individual protein via a common C-terminal dockerin (Doc) fusion tag. Genes of interest were chosen such that each gene product exhibited an identifiable unfolding pattern when loaded from the N to the C terminus. Each target protein was expressed with an N-terminal 11-amino-acid ybbR tag, which was used to covalently and site-specifically link the protein to the surface via Sfp synthase–catalyzed reaction with coenzyme A (CoA)[12]. At the C terminus the proteins contained a 75-amino-acid cellulosomal Doc from *Clostridium thermocellum*[13] as a specific handle targeted by the cohesin (Coh)-modified cantilever.

The gene microarray was aligned and reversibly bonded to a microfluidic chip known as MITOMI (mechanically induced trapping of molecular interactions). The chip has been used in the past for screening transcription factors[14,15] and mapping interaction networks[16]. More recently, our group employed MITOMI chips for molecular force assays[17]. In this work, MITOMI chips featured 640 dumbbell-shaped unit cells in a flow layer and 2,004 micromechanical valves in a control layer. Each unit cell was equipped with pneumatic 'neck', 'sandwich' and 'button' valves (**Fig. 1a**) according to design principles of soft lithography[18]. Each neck valve protected the microspotted DNA in the back chamber from exposure to other reagents during surface patterning in the front chamber. The sandwich valves prevented chamber-to-chamber cross contamination, ensuring that only a single protein variant was present in each sample spot. For surface chemistry in the front chamber, the button valves were actuated to shield the sample spots, allowing *n*-dodecyl β-D-maltoside passivation in the surrounding area. Releasing the button valves allowed subsequent functionalization with CoA-poly(ethylene glycol) (CoA-PEG) in the sample area under the buttons serving as the protein immobilization site. We expressed the genes by

[1]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-Universität, Munich, Germany. [2]Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität, Munich, Germany. [3]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, Munich, Germany. [4]These authors contributed equally to this work. Correspondence should be addressed to M.A.N. (michael.nash@lmu.de).

RECEIVED 3 MARCH; ACCEPTED 29 JULY; PUBLISHED ONLINE 7 SEPTEMBER 2014; CORRECTED AFTER PRINT 5 NOVEMBER 2014; DOI:10.1038/NMETH.3099

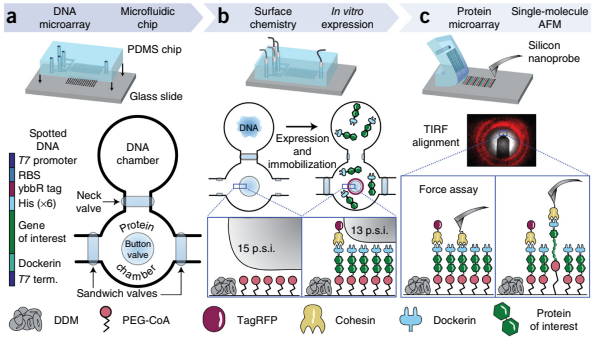


Figure 1 | Method workflow. (**a**) A gene array was spotted onto a glass slide. Genes were designed with a common set of flanking sequences, including a *T7* promoter region, ybbR tag, dockerin tag and *T7* terminator (term.). The multilayer microfluidic chip featuring 640 unit cells was aligned to the DNA microarray and bonded to the glass slide. Each unit cell comprised a DNA chamber, a protein chamber, and superseding elastomeric control valves actuated by pneumatic pressure. PDMS, poly(dimethylsiloxane). (**b**) Control valves were used for spatially selective surface modification of each protein chamber with poly(ethylene glycol)–coenzyme A (PEG-CoA) and for fluidic isolation of each chamber before *in vitro* expression of the microspotted DNA. Fluorescence labeling with TagRFP-cohesin was achieved by partial button-valve pressurization, leaving only an outer concentric ring of immobilized gene products exposed to the labeling solution. DDM, *n*-dodecyl β-D-maltoside. (**c**) After removal of the microfluidic device, the resulting well-defined, covalently attached protein microarray was accessed from above with a cohesin-functionalized atomic force microscope (AFM) cantilever. Single-molecule unfolding traces of each of the protein constructs were thus acquired sequentially at each corresponding array address with a single cantilever in a single experiment.

incubating an *in vitro* transcription and translation cell extract at 37 °C with the spotted DNA in the back chamber. The synthesized proteins then diffused to the front chamber, where they were covalently linked to the surface via an Sfp-catalyzed reaction of surface-bound CoA with solution-phase N-terminal ybbR peptide tags (**Fig. 1b**). Partial pressurization of the button valve[19] was used for tagging an outer concentric portion of the sample area with a fluorescently (TagRFP) tagged Coh that specifically bound to the C-terminal Doc tag of each target protein, thereby confirming successful protein synthesis and surface immobilization (**Supplementary Fig. 1**). Finally, the microfluidic device was removed from the glass slide to provide access to the protein array from above. Using this approach, we generated microarrays of site-specifically and covalently immobilized proteins for subsequent SMFS experiments, starting from a conventional gene array.

An inverted three-channel total-internal-reflection fluorescence/atomic force microscope (TIRF-AFM)[20] was used to position the cantilever in the center of the fluorescent rings in the protein array and perform SMFS measurements (**Fig. 1c**). The Coh-modified cantilever was used to probe the surface for expressed target proteins containing the C-terminal Doc tag. Upon surface contact of the cantilever, formation of a Coh-Doc complex allowed measurement of target-protein unfolding in a well-controlled pulling geometry (N to C terminus). We retracted the probe at constant velocity and recorded force-extension traces that characterized the unfolding fingerprint of the target protein. This approach-retract process could be repeated many times at each array address to characterize each expression construct.

Several unique features of the C-terminal Doc tag make it particularly suitable as a protein handle for SMFS. Its small size of 8 kDa does not notably add to the molecular weight of the gene products, which is advantageous for cell-free expression. Additionally, Doc exhibits a specific and high-affinity interaction with Coh domains from the *C. thermocellum* scaffold protein CipA. Coh was used both for fluorescence detection of the expression constructs and for modification of the cantilever. On the basis of our prior work, the Coh-Doc interaction is characterized to be high affinity, with a dissociation constant K_d in the low nanomolar range and rupture forces >125 pN at a loading rate of 10 nN/s (ref. 21). Our prior work also indicated that upon forced dissociation, Doc exhibited a characteristic double sawtooth rupture peak with a contour length increment of 8 nm separating the two peaks. We used this two-pronged double rupture event at the end of each force-extension trace as a positive indicator

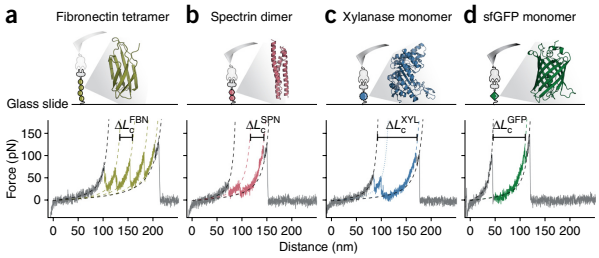


Figure 2 | Representative single-molecule force traces recorded in different protein spots on a single chip with a single cantilever. (**a**–**d**) Four proteins of interest, anchored between the coenzyme A (CoA)-functionalized surface and the cohesin-functionalized cantilever, were probed: fibronectin tetramer (**a**), spectrin dimer (**b**), xylanase monomer (**c**) and sfGFP monomer (**d**). The crystal structure and pulling configuration (top) are shown for each construct. Each single-molecule force-distance trace (bottom) shows the individual unfolding fingerprint of the respective protein of interest followed by a common, final double sawtooth peak (gray) that is characteristic of the cohesin-dockerin rupture. Experimental data were fitted with the worm-like chain model (dashed lines). Unfolding intermediates were also observed (fitted for only xylanase in **c**; dotted colored line).

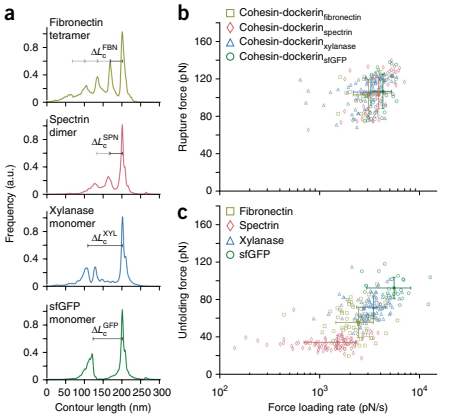


Figure 3 | Unfolding and rupture statistics from multiple force traces. (**a**) Relative frequency of observing given contour lengths determined by transforming and aligning multiple force traces into contour length space via the worm-like chain model. Shown are diagrams for the fibronectin tetramer (n = 27, ΔL_c^{FBN} = 33 nm), spectrin dimer (n = 50, ΔL_c^{SPN} = 34 nm), xylanase monomer (n = 91, ΔL_c^{XYL} = 93 nm) and sfGFP monomer (n = 25, ΔL_c^{GFP} = 79 nm). (**b**) Rupture force versus loading rate of the final cohesin-dockerin dissociation event. (**c**) Unfolding force versus loading rate for each protein of interest. The populations in **b** and **c** were fitted with two-dimensional Gaussians. Respective means and s.d. are plotted in the corresponding colors as solid symbols and error bars. a.u., arbitrary units.

that the gene of interest was completely expressed through to the C terminus (**Fig. 2**). Furthermore, this double rupture peak indicated that the interaction with the Coh-modified cantilever was specific and that the pulling geometry was strictly controlled such that force was applied to the molecule of interest from the N to the C terminus.

To validate and demonstrate our approach, we expressed genes of interest comprising well-known fingerprint domains in the SMFS literature. We produced multimeric polyproteins including tetrameric human type-III fibronectin (FBN)[22] and dimeric chicken brain α-spectrin (SPN)[23]. We also synthesized monomers of endo-1,4-xylanase T6 from *Geobacillus stearothermophilus* (XYL)[21], superfolder GFP (GFP)[24] and twitchin kinase[25]. In all cases, the target proteins were fused to N-terminal ybbR and C-terminal Doc tags (**Supplementary Figs. 2–6**). Unfolding data for FBN, SPN, XYL and GFP were obtained using a single cantilever to probe a single microarray (**Figs. 2** and **3**). Twitchin kinase was found not to express in sufficient yield to provide reliable unfolding statistics.

We transformed force-extension data (**Fig. 2**) into contour length space[26] using the worm-like chain model and compared the measured contour length increments with the amino acid sequence lengths of each protein and literature values. The observed contour lengths and rupture forces were consistent with our expectations. FBN showed a fourfold-repeated sequence of rupture peaks at contour length increments of 32 nm (ΔL_c^{FBN}; **Fig. 2a**) frequently interrupted by an intermediate peak at 10–12 nm, both features characteristic of FBN[22]. SPN showed two regular sawtooth-like peaks with contour lengths of 33 nm (ΔL_c^{SPN}; **Fig. 2b**)[23]. XYL exhibited a decreasing multipeaked unfolding fingerprint with a contour length increment of 92 nm (ΔL_c^{XYL}; **Fig. 2c**), occasionally showing additional increments corresponding to unfolding of remaining XYL subdomains, a result consistent with the prior study and accounting for N-terminal immobilization of XYL[21]. GFP unfolding showed a contour length increment of 74 nm (ΔL_c^{GFP}; **Fig. 2d**)[24]. As each protein in the array contained the same C-terminal Doc tag, the final two rupture peaks in all force traces represented rupture of the Coh-Doc complex regardless of the protein of interest.

In our system, surface densities of expressed proteins were comparable to those obtained in conventional SMFS experiments. Uninterpretable and nonspecific interactions were excluded from the analysis (**Supplementary Fig. 7**). By collecting multiple unfolding traces, we assembled contour length diagrams for each protein of interest[26,27] (**Fig. 3a**) and confirmed the predicted contour length increments on the basis of the encoded amino acid sequences in each DNA spot. Coh-Doc rupture events for all protein constructs in the array clustered to the same population in the force-loading rate plot, independently of the preceding rupture peaks from the protein of interest (**Fig. 3b**). The Coh-Doc ruptures agreed with previously reported values at similar loading rates[21]. The unfolding events of the proteins of interest produced distinct populations in the force-loading rate plots (**Fig. 3c**). The unfolding events depended on the internal structure and the unfolding pathway of the fingerprint domain when stretched between its N and C termini. SPN, for example, an elongated 3-helix bundle, was previously reported to exhibit a broader energy well (Δx = 1.7 nm; ref. 23) and showed a flatter distribution of unfolding forces than that of the more compact globular FBN domain with a shorter, steeper potential (Δx = 0.4 nm; ref. 22).

In summary, our flexible approach efficiently streamlines protein expression, purification and SMFS into a single integrated platform (**Supplementary Discussion**). The approach should be compatible with other *in vitro* expression systems including extracts derived from insects, rabbit reticulocytes and human cell lines, and it is capable of introducing post-translational modifications and non-natural amino acids, allowing, for example, the screening of site-directed mutants. Our method allows for synthesis of cytotoxic proteins or proteins with a tendency to form inclusion bodies during bulk expression. In addition to providing greatly improved throughput, our system has the advantage of measuring multiple constructs with one cantilever, thereby eliminating errors introduced when performing multiple calibrations on different samples with uncertainties of ~10% (ref. 28). Detecting subtle differences in mechanical stability with this high-throughput approach could therefore be used to perform mechanical phenotyping experiments on similarly stable families of mutant proteins. This workflow opens the door to large-scale screening studies of protein nanomechanical properties.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Addgene: pET28a-ybbR-HIS-sfGFP-DocI, 58708; pET28a-ybbR-HIS-CBM-CohI, 58709; pET28a-StrepII-TagRFP-CohI, 58710; pET28a-ybbR-HIS-Xyl-DocI, 58711;

pET28a-ybbR-HIS-10FNIII(x4)-DocI, 58712; pET28a-ybbR-HIS-Spec(x2)-DocI, 58713.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

M.O. is grateful to the Elite Network of Bavaria (IDK-NBT) for a doctoral fellowship. M.A.N. acknowledges support from Society in Science—The Branco Weiss Fellowship administered by the ETH Zürich. The authors acknowledge support from the DFG Sonderforschungsbereich 1032 and the European Research Council Grant Cellufuel (Advanced Grant 294438). The authors thank E. Bayer (Weizmann Institute) for starting genetic materials used for Doc and Coh modules.

AUTHOR CONTRIBUTIONS

M.O., M.A.N. and H.E.G. designed the research; M.O., W.O., M.A.J. and T.V. performed experiments; D.A.P. helped with immobilization strategies; M.O., W.O., M.A.J., L.F.M. and M.A.N. performed data analysis; M.O., W.O., M.A.J., M.A.N. and H.E.G. cowrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Thomas, W.E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E.V. *Cell* **109**, 913–923 (2002).
2. Li, C. & Xu, Q. *Cell. Signal.* **12**, 435–445 (2000).
3. Engler, A.J., Sen, S., Sweeney, H.L. & Discher, D.E. *Cell* **126**, 677–689 (2006).
4. Müller, D., Helenius, J., Alsteens, D. & Dufrêne, Y.F. *Nat. Chem. Biol.* **5**, 383–390 (2009).
5. Florin, E.-L., Moy, V.T. & Gaub, H.E. *Science* **264**, 415–417 (1994).
6. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. & Gaub, H. *Science* **276**, 1109–1112 (1997).
7. Fernandez, J.M. & Li, H. *Science* **303**, 1674–1678 (2004).
8. Oesterhelt, F. *et al. Science* **288**, 143–146 (2000).
9. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. & Block, S.M. *Nature* **438**, 460–465 (2005).
10. Bryant, Z. *et al. Nature* **424**, 338–341 (2003).
11. Linke, W.A. *Cardiovasc. Res.* **77**, 637–648 (2008).
12. Yin, J. *et al. Proc. Natl. Acad. Sci. USA* **102**, 15815–15820 (2005).
13. Bayer, E.A., Belaich, J.-P., Shoham, Y. & Lamed, R. *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
14. Maerkl, S.J. & Quake, S.R. *Science* **315**, 233–237 (2007).
15. Rockel, S., Geertz, M., Hens, K., Deplancke, B. & Maerkl, S.J. *Nucleic Acids Res.* **41**, e52 (2013).
16. Gerber, D., Maerkl, S.J. & Quake, S.R. *Nat. Methods* **6**, 71–74 (2009).
17. Otten, M., Wolf, P. & Gaub, H.E. *Lab Chip* **13**, 4198–4204 (2013).
18. Thorsen, T., Maerkl, S.J. & Quake, S.R. *Science* **298**, 580–584 (2002).
19. Garcia-Cordero, J.L. & Maerkl, S.J. *Chem. Commun.* **49**, 1264–1266 (2013).
20. Gumpp, H., Stahl, S.W., Strackharn, M., Puchner, E.M. & Gaub, H.E. *Rev. Sci. Instrum.* **80**, 063704 (2009).
21. Stahl, S.W. *et al. Proc. Natl. Acad. Sci. USA* **109**, 20431–20436 (2012).
22. Li, L., Huang, H.H.-L., Badilla, C.L. & Fernandez, J.M. *J. Mol. Biol.* **345**, 817–826 (2005).
23. Rief, M., Pascual, J., Saraste, M. & Gaub, H.E. *J. Mol. Biol.* **286**, 553–561 (1999).
24. Dietz, H. & Rief, M. *Proc. Natl. Acad. Sci. USA* **103**, 1244–1247 (2006).
25. Greene, D.N. *et al. Biophys. J.* **95**, 1360–1370 (2008).
26. Puchner, E.M., Franzen, G., Gautel, M. & Gaub, H.E. *Biophys. J.* **95**, 426–434 (2008).
27. Jobst, M.A., Schoeler, C., Malinowska, K. & Nash, M.A. *J. Vis. Exp.* **82**, e50950 (2013).
28. Gibson, C.T., Smith, D.A. & Roberts, C.J. *Nanotechnology* **16**, 234–238 (2005).

ONLINE METHODS

Chip fabrication. Ready-to-use wafers for flow and control layers of the 640-chamber MITOMI design were obtained from Stanford Microfluidics Foundry (design name DTPAd)[14]. The flow wafer features 15-µm-high features, rounded by photoresist reflow, whereas the control wafer features a rectangular cross-section.

Microfluidic chips were cast in poly(dimethylsiloxane) (PDMS) from these wafers. For the control layer, Sylgard 184 (Dow Corning) base and curing agent were mixed at a ratio of 5:1 by weight, poured onto the wafer, degassed and partially cured for 20 min at 80 °C. For the flow-layer wafer, a 20:1 base–to–curing agent mixture of Sylgard 184 was spin-coated for 75 s at 1,600 r.p.m. and partially cured for 30 min at 80 °C. The control layer chips were cut out, inlet holes were punched and the chips were aligned onto the spin-coated PDMS on the flow-layer wafer. After the two-layer chips were baked for 90 min at 80 °C, they were cut and removed from the wafer, and inlet/outlet holes were punched. Microfluidic chips were stored for up to 6 weeks.

Cloning. For the construction of the fusion proteins, Gibson assembly[29] was used. A ratio of 0.07 pmol vector to 0.3 pmol of insert was used for the fusion reaction. The primer sequences are provided in **Supplementary Table 1**. A pET28a plasmid was linearized with primers 1 and 2. The dockerin type I–encoding gene was isolated from the xylanase-dockerin type I construct[21] with primers 3 and 4. Codon-optimized sequences were purchased from GeneArt/Invitrogen. The genes of interest were designed in such a way that they already contained sequences overlapping those of their neighboring partners (pET28a and dockerin type I). In the case of the spectrin, two domains were linked with a flexible glycine-serine (×6) linker. For fibronectin, four type III domains were fused separated by glycine-serine (×6) linkers. The expression vector in all cases was a pET28a plasmid with a modified multiple cloning site (sequence attached: plasmids are available at Addgene, **Supplementary Table 2**). After construction, clones were verified via sequencing and amplified in NEB 5-alpha *Escherichia coli* cells. Following plasmid preparation, samples were concentrated up to 500 ng/µl before microspotting.

DNA microspotting. A 24 × 60–mm #1 thickness coverslip (Thermo Scientific) was silanized with 3-aminopropyldimethylethoxysilane (ABCR) following literature protocols[30].

The DNA solution containing 1% (w/v) nuclease-free bovine serum albumin (Carl Roth) in nuclease-free water was microspotted under humid atmosphere onto the silanized coverslip using the GIX Microplotter II (Sonoplot) and a glass capillary with a 30-µm tip diameter (World Precision Instruments) according to the manufacturer's instructions in a rectangular 40 × 16 pattern with 320-µm column pitch and 678-µm row pitch. Alignment of the DNA array and the microfluidic chip was done manually using a stereomicroscope. Bonding between the glass cover slip and microfluidic device was achieved by thermal bonding for 5 h at 80 °C on a hot plate.

Protein synthesis on-chip. The microfluidic device was operated at a pressure of 4 p.s.i. in the flow layer and 15 p.s.i. in the control layer. Operation started with the button and neck valves actuated for surface passivation. The flow layer was passivated by flushing through standard buffer (25 mM Tris, 75 mM NaCl, 1 mM CaCl2, pH 7.2) for 5 min and 2% *n*-dodecyl β-D-maltoside (Thermo Scientific) in nuclease-free H_2O for 30 min (ref. 31). Next the button valve was opened, and borate buffer (50 mM sodium borate, pH 8.5) was flushed through for 30 min to deprotonate aminosilane groups on the glass surface.

For maleimide/coenzyme A functionalization, a solution of 5 mM NHS-PEG-maleimide (MW = 513 Da, Thermo Scientific) in borate buffer was flushed through for 45 min. The device was then rinsed with nuclease-free H_2O for 5 min, followed by 30 min of 20 mM coenzyme A (Merck) in coupling buffer (50 mM sodium phosphate, pH 7.2, 50 mM NaCl, 10 mM EDTA). The button valve was then actuated to protect the functionalized surface area followed by 5 min of rinsing with standard buffer.

S30 T7 HY (Promega) *in vitro* transcription and translation mix, supplemented with 1 µL T7 polymerase (Promega) and 0.5 µL RNase inhibitor (Invitrogen), was then flushed into the chip, filling the DNA chambers (neck valve open).

The neck valve was then closed, and the channels were filled with 4′-phosphopantetheinyl transferase (Sfp synthase) in Sfp buffer (50 mM HEPES, 10 mM $MgCl_2$). The chip was then incubated at 37 °C on a hot plate. After 1 h of incubation, the neck and the button valves were opened to allow Sfp synthase–catalyzed linkage of expressed protein to the coenzyme A–functionalized area below the button. At the same time the sandwich valves were actuated to avoid chamber-to-chamber cross-contamination. After another 1.5 h of incubation, the neck and button valves were closed, the sandwich valves were opened and the chip was rinsed with standard buffer for 20 min.

To verify successful protein expression and immobilization on the functionalized surface area, a fluorescent detection construct (TagRFP–cohesin type I (2 µg/ml) in standard buffer) was flushed through the device for 10 min with the button valve actuated. The sandwich valves were then actuated, and the button valve partially released by decreasing the pressure to 11 p.s.i. After 20 min of incubation at room temperature, the sandwich valves were opened, and the chip flushed with standard buffer for 20 min. Fluorescence images of all chambers were recorded on an inverted microscope with a 10× objective (Carl Zeiss), featuring an electron-multiplying charge-coupled device (EMCCD) camera (Andor). Prior to force spectroscopy experiments, the chip was stored in buffer at 4 °C.

Directly before measurement, the PDMS chip was peeled off from the glass substrate under buffer, revealing the microarray while avoiding drying of the functionalized surface. The array surface was then rinsed several times with buffer. We did not encounter any problems with cross-contamination between chambers.

Cantilever functionalization. A silicon-nitride cantilever bearing a silicon tip with a tip radius of ~8 nm (Biolever mini, Olympus) was silanized with ABCR as described previously[30]. Protein functionalization was performed in a similar way as reported previously[27,31]. Briefly, a 50 µM solution of CBM A2C–cohesin from *C. thermocellum* in standard buffer was incubated with 1:2 (v/v) TCEP beads (Tris (2-carboxyethyl) phosphine disulfide reducing gel, Thermo Scientific), previously washed with standard buffer, for 2.5 h. The cantilever was submerged in borate buffer for 45 min to deprotonate primary amine groups on the silanized surface and then incubated with 20 mM NHS-PEG-maleimide (MW = 5 kDa, Rapp Polymere) in borate buffer for 60 min.

doi:10.1038/nmeth.3099

The cantilever was rinsed sequentially in three beakers of deionized H_2O. TCEP beads were separated from the protein solution by centrifugation at 1,000*g* for 1 min. Next the cantilever was incubated for 60 min with reduced protein solution, which was diluted to a concentration of 1 mg/mL with standard buffer. Finally the cantilever was rinsed sequentially in three beakers of standard buffer and stored submerged in standard buffer in humid atmosphere at 4 °C for up to 24 h before use.

Force spectroscopy. A custom-built TIRF (total internal reflection fluorescence)-AFM (atomic force microscope) hybrid[20,30] was used to conduct the force spectroscopy measurements. The TIRF microscope was used to image fluorophores in up to three different color channels simultaneously using an iChrome MLE-S four-color laser (Toptica Photonics), an Optosplit III triple emission image splitter (Cairn Research) and a Xion3 EMCCD camera (Andor). A long-range stick-slip *xy* piezo nanopositioning system (ANC350, Attocube Systems) allowed access to the whole microchip array as well as fine spatial sampling of different surface molecules on the nanometer scale within each protein spot. Cantilever actuation in the *z* direction was performed by a LISA piezo-actuator (Physik Instrumente) driven by an MFP3D AFM controller (Asylum Research).

The following force spectroscopy protocol was performed repeatedly in each functionalized protein target area. The cantilever approach velocity was 3,000 nm/s, dwell time at the surface was 10 ms and retract velocity was 800 nm/s. Data were recorded with 6,250-Hz sampling rate. The cantilever typically had a spring constant in the range of 100 pN/nm and a resonance frequency of 25 kHz in water. Accurate calibration of the system was performed by the nondestructive thermal method[32,33] using corrections to account for discrepancies from the original theory[27,34].

Data and statistical analysis. The raw data were converted from photodiode voltages into force values in newtons, and the following standard corrections were applied. The zero force value for the unloaded cantilever in each curve was determined by averaging over 40-nm extension after the final complex rupture and subtracting this value from each force value in the curve. The position of the surface was determined by finding the force value closest to 0 in a small neighborhood of the first non-negative force value in the force-extension trace. The *z* piezo position was corrected for the true tip-sample separation due to deflection of the lever as a function of the force for a Hookean spring.

A pattern-recognition software based on a package described previously[26] and adapted in-house chose the curves showing worm-like chain force responses of the stretched protein constructs. Example curves showing multiple, unspecific or no interactions are shown in **Supplementary Figure 7**, together with a single xylanase trace for comparison. The expected protein backbone contour length increments for each construct were detected in contour length space: the real part of the following numerically solved inverse worm-like chain (WLC) formula[27] was used to transform force-extension data into force–contour length space for every measured force curve:

$$L(x) = \frac{x}{6u}\left(3 + 4u + \frac{9 - 3u + 4u^2}{g(u)} + g(u)\right)$$

where

$$g(u) = \left(27 - \frac{27}{2}u + 36u^2 - 8u^3 + \frac{3\sqrt{3}}{2}\sqrt{-u^2\left((4u-3)^3 - 108\right)}\right)^{\frac{1}{3}}$$

and

$$u = F\frac{L_p}{kT}$$

with L the contour length, x the extension, F the force, L_p the persistence length, k Boltzmann's constant and T the temperature. Transformed data points were combined in a Gaussian kernel density estimate with a bandwidth of 1 nm and plotted with a resolution of 1 nm. In these resulting energy-barrier position diagrams, the contour length increments could easily be determined. The transformation was performed with the following parameters: persistence length $L_p = 0.4$ nm, thermal energy $kT = 4.1$ pN nm. Force and distance thresholds were applied at 10 pN and 5 nm, respectively. The measurement data sets in each protein spot on the chip typically showed a yield of 0.5–5% specific interactions.

The force peaks corresponding to protein domain unfolding events, as well as those corresponding to final ruptures, were line fitted in force-time space to measure the loading rate of each individual event.

WLC fits for demonstrative purposes in **Figure 2** were done by using the following formula:

$$F(x) = \frac{kT}{L_p}\left(\frac{1}{4(1 - x/L)^2} + \frac{x}{L} - \frac{1}{4}\right)$$

with F the force, k the Boltzmann's constant, T the temperature, L_p the persistence length, x the extension and L the contour length.

Discrepancies between contour length increments in fitted single-molecule traces and aligned contour length diagrams are artifacts caused by the fixed persistence length in the contour length transformation, whereas the WLC fits to single force traces treat both contour length and persistence length of each stretch as free parameters. An overview of the yield of interpretable curves of all constructs is available in **Supplementary Table 3**.

29. Gibson, D.G. *et al. Nat. Methods* **6**, 343–345 (2009).
30. Zimmermann, J.L., Nicolaus, T., Neuert, G. & Blank, K. *Nat. Protoc.* **5**, 975–985 (2010).
31. Huang, B., Wu, H., Kim, S. & Zare, R.N. *Lab Chip* **5**, 1005–1007 (2005).
32. Hutter, J.L. & Bechhoefer, J. *Rev. Sci. Instrum.* **64**, 1868 (1993).
33. Cook, S.M. *et al. Nanotechnology* **17**, 2135–2145 (2006).
34. Proksch, R., Schäffer, T.E., Cleveland, J.P., Callahan, R.C. & Viani, M.B. *Nanotechnology* **15**, 1344–1350 (2004).

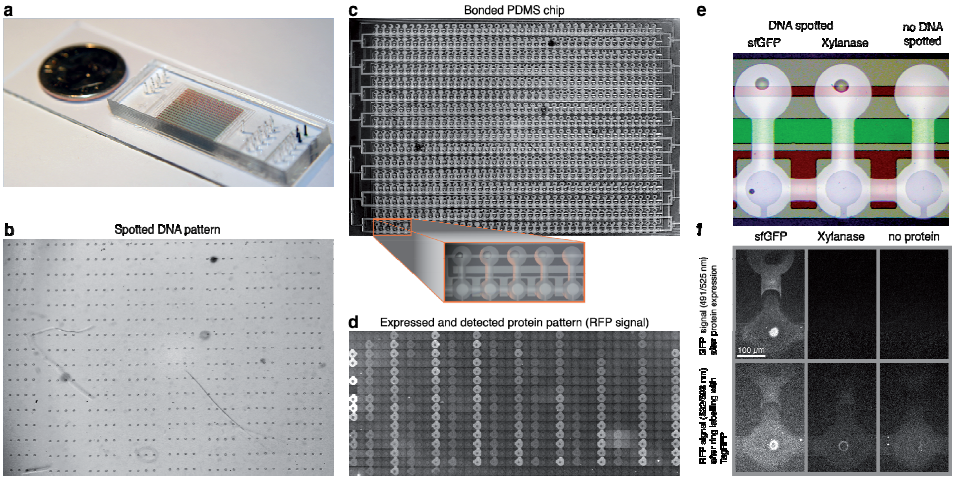
doi:10.1038/nmeth.3099

CORRIGENDA

Corrigendum: From genes to protein mechanics on a chip

Marcus Otten, Wolfgang Ott, Markus A Jobst, Lukas F Milles, Tobias Verdorfer, Diana A Pippig, Michael A Nash & Hermann E Gaub
Nat. Methods **11, 1127–1130 (2014); published online 7 September 2014; corrected after print 5 November 2014**

In the version of this article initially published, the grant "European Research Council Grant Cellufuel (Advanced Grant 294438)" was mistakenly left out of the Acknowledgements. The error has been corrected in the HTML and PDF versions of the article.

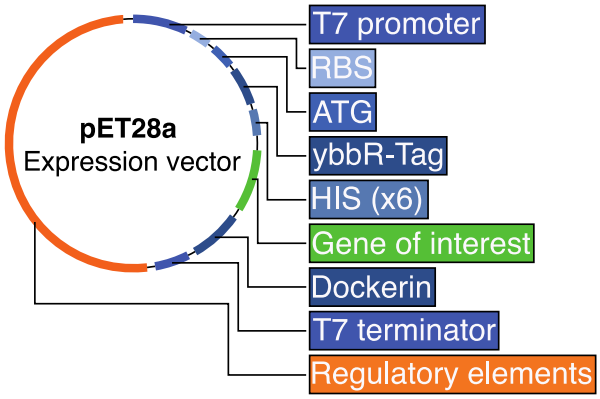


Supplementary Figure 1

Microfluidic chip overview.

(a) Photograph of a microfluidic chip bonded to a glass slide with a US dime for scale. Control channels are filled with food dye for better visualization. (b) Pattern of a typical DNA array, consisting of repeats of rows with four different genes and one row with nothing spotted as negative control. (c) Photograph of a bonded PDMS chip onto the glass slide with DNA spots in the back chamber. The orange highlighted frame shows a zoom in of the bottom left corner. (d) Typical fluorescence collage assembled from 640 single fluorescence micrographs of each protein chamber on one single chip shows pattern of expressed protein (assembly not to scale). Fluorescence signal of TagRFP reveals expression levels and Dockerin specificity. Here, low passivation of the protein chamber facilitates visualization. (e) Three of 640 adjacent dumbbell-shaped chambers, one with sfGFP DNA spotted (left), one with Xylanase DNA (center) and one negative control without DNA (right). Control channels are visualized with food dye: neck valve (green), sandwich valve (red), and button valve (blue). (f) Fluorescence images showing GFP signal (top) from expressed and immobilized ybbR-sfGFP-Dockerin (left), ybbR-Xylanase-Dockerin (center) with negative control lacking the spotted DNA (right). The bottom row shows the signal from the TagRFP detection construct, which specifically bound to the Dockerin tag via the Cohesin domain.

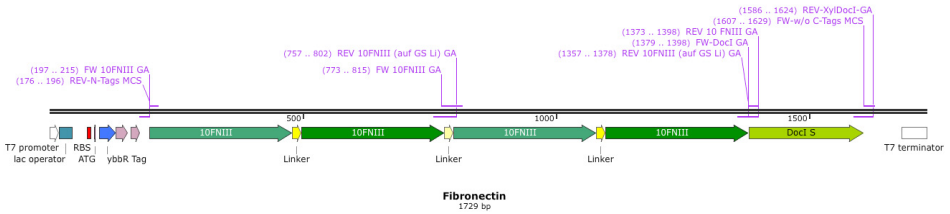
Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 2

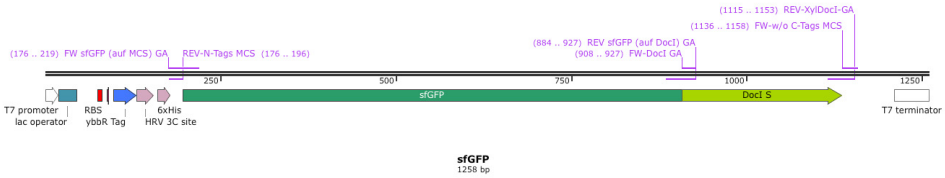
Diagram of the expression vector pET28a with an individual gene of interest.

Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 3

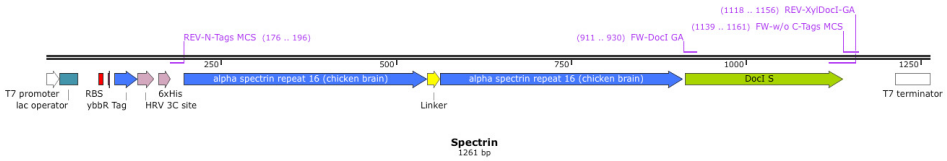
Schematic of the fibronectin tetramer gene cassette.



Supplementary Figure 4

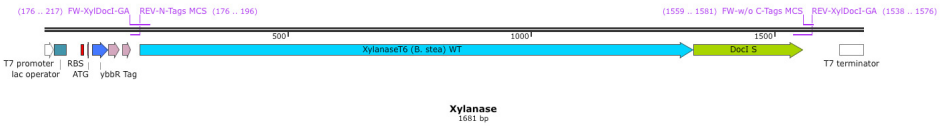
Schematic of the sfGFP dimer gene cassette.

Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 5

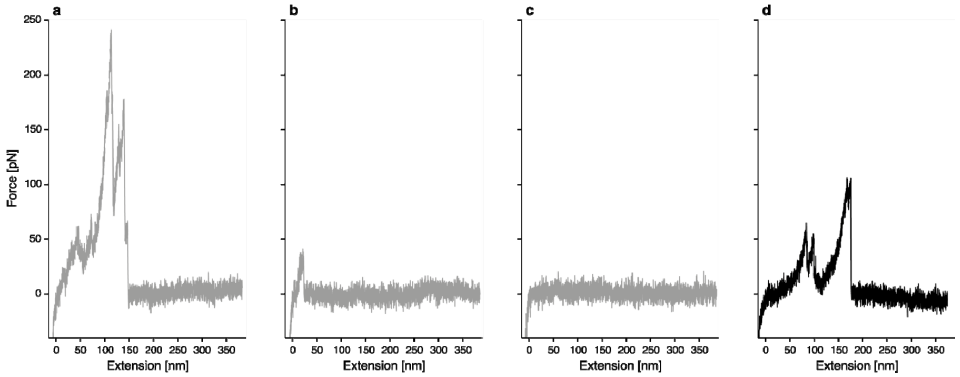
Schematic of the spectrin dimer gene cassette.



Supplementary Figure 6

Schematic of the xylanase gene cassette.

Nature Methods: doi:10.1038/nmeth.3099



Supplementary Figure 7

Exemplary force traces

Example curves showing (a) uninterpretable interaction, (b) non-specific interaction of cantilever with surface, (c) no interaction, and (d) a specific Xylanase-Dockerin unfolding and unbinding trace. Curves similar to those shown in a-c were excluded from the analysis.

Nature Methods: doi:10.1038/nmeth.3099

Supplementary Discussion

Typically in SMFS experiments, rupture force – loading rate plots are used to characterize k_{off} and Δx, the unbinding (or unfolding) probability per time unit and the distance to the transition state along the reaction coordinate, respectively, providing direct information about the energy landscape governing protein folding[1]. SMFS experiments are also complemented by all-atom simulations of such systems *in silico*. Recently, it was shown that high speed SMFS experiments could be performed at speeds achievable in molecular dynamics simulations[2], overcoming a long standing discrepancy between experiment and simulation.

In analyzing single-molecule unfolding curves (i.e., **Fig. 2**), we note that the spotted DNA at the measured array addresses correctly corresponded to the domain of interest encoded by the corresponding spotted DNA at that position. For example, the fibronectin tetramer was measured at array position (237), the spectrin dimer at position (239), the xylanase monomer at position (196), and the sfGFP monomer at position (238), corresponding to the correct genes deposited into the expression chambers at those array positions (**Fig. 2**). Typically 10–15 immobilization chambers per microarray were measured. Typically several thousand force curves were acquired giving rise to dozens of interpretable single-molecule interaction curves.

Upper force limit

Here we extend the discussion regarding the upper force limit for the SMFS-MITOMI system. In all force-distance data traces, the last rupture events represent unbinding of the Coh-Doc complex, not unfolding of a domain. This rupture force of the Coh-Doc complex represents an upper limit in force for the entire construct, since the Doc is used as a handle sequence grabbed by the Coh-modified cantilver. The system we described can therefore interrogate domains with mechanical rupture forces that lie below that of Coh-Doc (~125 pN at 10 nN/s). If proteins with larger unfolding forces should be investigated, other Coh-Doc domains that show even higher complex rupture forces can be used. The Coh-Doc pair from *R. flavefaciens*, for example (PDB 4IU3) exhibits rupture forces over 600 pN at these loading rates (unpublished data). This could alternatively be used as a handle sequence to interrogate mechanically more stable domains of interest.

Computerized image analysis can be used to automate cantilever positioning above the fluorescent rings and subsequent acquisition of unfolding traces at each array address in combination with online force curve analysis to further increase throughput. Additionally, well-characterized reference proteins on the same chip may serve as calibration standards further minimizing uncertainty in absolute force values.

It is possible to operate the MITOMI device in a simplified way without the need for microspotting template DNA and chip alignment. This manual option should encourage the interested community to apply the suggested method to their single molecule force spectroscopy experiments. MITOMI enables the experimenter to prepare up to 16 different constructs in one column with 40 repeats each by flow-loading the DNA. Since

Nature Methods: doi:10.1038/nmeth.3099

the valves are pressure sensitive it is also possible to operate these manually. This way it is possible to make use of the parallelized method without having the automation tools.Supplementary Materials & Methods

DNA Sequences

Supplementary Table 1. Overview of primers

	Name	Sequence
1	FW-w/o C-Tags MCS	TAACTCGAGTAAGATCCGGCTGC
2	REV-N-Tags MCS	GCTAGCACTAGTCCATGGGTG
3	FW-DocI GA	AAAGTGGTACCTGGTACTCC
4	REV-XylDocI-GA	CGGATCTTACTCGAGTTAGTTCTTGTACGGCAATGTATC
5	FW 10FNIII GA	CGCACCGGCTCTGGCTCTGGCTCTGTTAGTGATGTTCCGCGTG
6	REV 10 FNIII GA	GGAGTACCAGGTACCACTTTGGTGCG
7	REV 10FNIII (auf GS Li) GA	ACTAACAGAGCCAGAGCCAGAGCCGGTGCGATAATTGATTGAAATC
8	FW sfGFP (auf MCS) GA	CACCCATGGACTAGTGCTAGCAGCAAAGGTGAAGAACTGTTTAC
9	REV sfGFP (auf DocI) GA	GGAGTACCAGGTACCACTTTCTTATACAGCTCATCCATACCATG

Supplementary Table 2. Overview of DNA plasmids available at Addgene database

Addgene ID	Construct
58708	pET28a-ybbR-HIS-sfGFP-DocI
58709	pET28a-ybbR-HIS-CBM-CohI
58710	pET28a-StrepII-TagRFP-CohI
58711	pET28a-ybbR-HIS-Xyl-DocI
58712	pET28a-ybbR-HIS-10FNIII(x4)-DocI
58713	pET28a-ybbR-HIS-Spec(x2)-DocI

Multiple cloning site for the protein of interest:

N terminal region

T7 promoter | *lac operator* | *RBS* | *ATG* | *ybbr Tag* | *HRV 3C protease site* | *HIS Tag (x6)*

TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|GACTCTCTGGAATTCATCGCTTCTAA
ACTGGCT|CTGGAAGTTCTGTTCCAGGGTCCG|CTGCAG|CACCACCACCACCACCAC|CCATGG
ACTAGTGCTAGC

C terminal region

Dockerin Type I | T7 terminator

AAAGTGGTACCTGGTACTCCTTCTACTAAATTATACGGCGACGTCAATGATGACGGAAAAGTTAA
CTCAACTGACGCTGTAGCATTGAAGAGATATGTTTTGAGATCAGGTATAAGCATCAACACTGACA
ATGCCGATTTGAATGAAGACGGCAGAGTTAATTCAACTGACTTAGGAATTTTGAAGAGATATATT
CTCAAAGAAATAGATACATTGCCGTACAAGAAC|TAA|CTCGAGTAAGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTT

10 FibronectinIII (4x):

Glycin-Serin Linker (x6)

GTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAGCCTGCTGATTAGCTG
GGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAACCGGTGGTAATAGTC
CGGTTCAAGAATTTACCGTTCCGGGTAGCAAAAGCACCGCAACCATTAGCGGTCTGAAACCGGGT
GTTGATTACACCATTACCGTTTATGCCGTTACCGGTCGTGGTGATTCACCGGCAAGCAGCAAACC
GATTAGCATTAACTATCGTACCGGTAGCGGTAGTGGTAGCGTTTCAGATGTGCCTCGCGACCTGG
AAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCCTGCCGTGACCGTGCGC
TATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGAATTCACAGTTCCTGG
TTCAAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACACAATCACAGTGTATG
CAGTGACAGGTCGCGGTGATAGTCCGGCAAGTTCAAAACCGATTTCAATCAATTATCGCACCGGC
TCTGGCTCTGGCTCTGTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAG
CCTGCTGATTAGCTGGGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAA
CCGGTGGTAATAGTCCGGTTCAAGAATTTACCGTTCCGGGTAGCAAAAGCACCGCAACCATTAGC
GGTCTGAAACCGGGTGTTGATTACACCATTACCGTTTATGCCGTTACCGGTCGTGGTGATTCACC
GGCAAGCAGCAAACCGATTAGCATTAACTATCGTACCGGTAGCGGTAGTGGTAGCGTTTCAGATG
TGCCTCGCGACCTGGAAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCCT
GCCGTGACCGTGCGCTATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGA
ATTCACAGTTCCTGGTTCAAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACA
CAATCACAGTGTATGCAGTGACAGGTCGCGGTGATAGTCCGGCAAGTTCAAAACCGATTTCAATC
AAttaTCGCACC

Nature Methods: doi:10.1038/nmeth.3099

sfGFP:

AGCAAAGGTGAAGAACTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGTGATGTTAA
TGGCCACAAATTTTCAGTTCGTGGTGAAGGCGAAGGTGATGCAACCATTGGTAAACTGACCCTGA
AATTTATCTGTACCACCGGCAAACTGCCGGTTCCGTGGCCGACCCTGGTTACCACCCTGACCTAT
GGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATGAAACGCCACGATTTTTTCAAAAGCGCAAT
GCCGGAAGGTTATGTTCAAGAACGTACCATCTCCTTTAAAGACGACGGTAAATACAAAACCCGTG
CCGTTGTTAAATTTGAAGGTGATACCCTGGTGAATCGCATTGAACTGAAAGGCACCGATTTTAAA
GAGGATGGTAATATCCTGGGCCACAAACTGGAATATAATTTCAATAGCCACAACGTGTATATCAC
CGCAGACAAACAGAAAAATGGCATCAAAGCCAATTTTACCGTGCGCCATAATGTTGAAGATGGTA
GCGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTTCTGCTGCCG
GATAATCATTATCTGAGCACCCAGACCGTTCTGAGCAAAGATCCGAATGAAAAACGTGATCATAT
GGTGCTGCATGAGTATGTTAATGCAGCAGGTATTACCCATGGTATGGATGAGCTGTATAAG

alpha-Spectrin repeat 16 (chicken brain) (x2):

Glycin-Serine Linker (x6)

CGTGCTAAACTGAACGAATCTCACCGTCTGCACCAGTTCTTCCGTGACATGGACGACGAAGAATC
TTGGATCAAAGAAAAAAAACTGCTGGTTTCTTCTGAAGACTACGGTCGTGACCTGACCGGTGTTC
AGAACCTGCGTAAAAAACACAAACGTCTGGAAGCTGAACTGGCTGCTCACGAACCGGCTATCCAG
GGTGTTCTGGACACCGGTAAAAAACTGTCTGACGACAACACCATCGGTAAAGAAGAAATCCAGCA
GCGTCTGGCTCAGTTCGTTGACCACTGGAAAGAACTGAAACAGCTGGCTGCTGCTCGTGGTCAGC
GTCTGGAAGAATCTCTGGAATACGGTAGCGGTAGCGGTTCACGTGCTAAACTGAACGAATCTCAC
CGTCTGCACCAGTTCTTCCGTGACATGGACGACGAAGAATCTTGGATCAAAGAAAAAAAACTGCT
GGTTTCTTCTGAAGACTACGGTCGTGACCTGACCGGTGTTCAGAACCTGCGTAAAAAACACAAAC
GTCTGGAAGCTGAACTGGCTGCTCACGAACCGGCTATCCAGGGTGTTCTGGACACCGGTAAAAAA
CTGTCTGACGACAACACCATCGGTAAAGAAGAAATCCAGCAGCGTCTGGCTCAGTTCGTTGACCA
CTGGAAAGAACTGAAACAGCTGGCTGCTGCTCGTGGTCAGCGTCTGGAAGAATCTCTGGAATAt

Xylanase:

AAGAATGCAGATTCCTATGCGAAAAAACCTCACATCAGCGCATTGAATGCCCCACAATTGGATCA
ACGCTACAAAAACGAGTTCACGATTGGTGCGGCAGTAGAACCTTATCAACTACAAAATGAAAAAG
ACGTACAAATGCTAAAGCGCCACTTCAACAGCATTGTTGCCGAGAACGTAATGAAACCGATCAGC
ATTCAACCTGAGGAAGGAAAATTCAATTTTGAACAAGCGGATCGAATTGTGAAGTTCGCTAAGGC
AAATGGCATGGATATTCGCTTCCATACACTCGTTTGGCACAGCCAAGTACCTCAATGGTTCTTTC
TTGACAAGGAAGGTAAGCCAATGGTTAATGAATGCGATCCAGTGAAACGTGAACAAAATAAACAA
CTGCTGTTAAAACGACTTGAAACTCATATTAAAACGATCGTCGAGCGGTACAAAGATGACATTAA
GTACTGGGACGTTGTAAATGAGGTTGTGGGGGACGACGGAAAACTGCGCAACTCTCCATGGTATC
AAATCGCCGGCATCGATTATATTAAAGTGGCATTCCAAGCAGCTAGAAAATATGGCGGAGACAAC
ATTAAGCTTTACATGAATGATTACAATACAGAAGTCGAACCGAAGCGAACCGCTCTTTACAATTT
AGTCAAACAACTGAAAGAAGAGGGTGTTCCGATCGACGGCATCGGCCATCAATCCCACATCCAAA
TCGGCTGGCCTTCTGAAGCAGAAATCGAGAAAACGATTAACATGTTCGCCGCTCTCGGTTTAGAC
AACCAAATCACTGAGCTTGATGTGAGCATGTACGGTTGGCCGCCGCGCGCTTACCCGACGTATGA
CGCCATTCCAAAACAAAAGTTTTTGGATCAGGCAGCGCGCTATGATCGTTTGTTCAAACTGTATG
AAAAGTTGAGCGATAAAATTAGCAACGTCACCTTCTGGGGCATCGCCGACAATCATACGTGGCTC
GACAGCCGTGCGGATGTGTACTATGACGCCAACGGGAATGTTGTGGTTGACCCGAACGCTCCGTA
CGCAAAAGTGGAAAAAGGGAAAGGAAAAGATGCGCCGTTCGTTTTTGGACCGGATTACAAAGTCA
AACCCGCATATTGGGCTATTATCGACCAC

Nature Methods: doi:10.1038/nmeth.3099

Detection construct RFP-Cohesin:

TagRFP-Cohesin:

T7 promoter | *lac operator* | *RBS* | *ATG* | *StrepII Tag* | *TagRFP* | *Linker* | *Cohesin* | T7 terminator

TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|TGGTCTCACCCGCAGTTCGAAAAA|G
TTTCTAAAGGTGAAGAACTGATCAAAGAAAACATGCACATGAAACTGTACATGGAAGGTACTGTT
AACAACCACCACTTCAAATGCACCTCTGAAGGTGAAGGTAAACCGTACGAAGGTACTCAGACCAT
GCGTATCAAAGTTGTTGAAGGTGGTCCGCTGCCGTTCGCTTTCGACATCCTGGCTACCTCTTTCA
TGTACGGTTCTCGTACCTTCATCAACCACACCCAGGGTATCCCGGACTTCTTCAAACAGTCTTTC
CCGGAAGGTTTCACCTGGGAACGTGTTACCACCTACGAAGACGGTGGTGTTCTGACCGCTACCCA
GGACACCTCTCTGCAAGACGGTTGCCTGATCTACAACGTTAAAATCCGTGGTGTTAACTTCCCGT
CTAACGGTCCGGTTATGCAGAAAAAAACCCTGGGTTGGGAAGCTAACACCGAAATGCTGTACCCG
GCTGACGGTGGTCTGGAAGGTCGTTCTGACATGGCTCTGAAACTGGTTGGTGGTGGTCACCTGAT
CTGCAACTTCAAAACCACCTACCGTTCTAAAAAACCGGCTAAAAACCTGAAAATGCCGGGTGTTT
ACTACGTTGACCACCGTCTGGAACGTATCAAAGAAGCTGACAAAGAAACCTACGTTGAACAGCAC
GAAGTTGCTGTTGCTCGTTACTGCGACCTGCCGTCTAAACTGGGTCACAAACTGAAC|GGCAGTG
TAGTACCATCAACACAGCCTGTAACAACACCACCTGCAACAACAAAACCACCTGCAACAACAATA
CCGCCGTCAGATGATCCGAATGCA|GGATCCGACGGTGTGGTAGTAGAAATTGGCAAAGTTACGG
GATCTGTTGGAACTACAGTTGAAATACCTGTATATTTCAGAGGAGTTCCATCCAAAGGAATAGCA
AACTGCGACTTTGTGTTCAGATATGATCCGAATGTATTGGAAATTATAGGGATAGATCCCGGAGA
CATAATAGTTGACCCGAATCCTACCAAGAGCTTTGATACTGCAATATATCCTGACAGAAAGATAA
TAGTATTCCTGTTTGCGGAAGACAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTT
GCAAAAATAAGAGCAACTGTAAAATCAAGTGCTCCGGGCTATATTACTTTCGACGAAGTAGGTGG
ATTTGCAGATAATGACCTGGTAGAACAGAAGGTATCATTTATAGACGGTGGTGTTAACGTTGGCA
ATGCAACA|TAA|CTCGAGTAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG
CTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT
TTT

Molecular weights of synthesized fusion proteins

ybbR-(Fibronectin)$_4$-Dockerin Type I: 53 kDa
ybbR-(Spectrin)$_2$-Dockerin Type I: 40 kDa
ybbR-Xylanase-Dockerin Type I: 56 kDa
ybbR-sfGFP-Dockerin Type I: 39 kDa
ybbR-Twitchin-Dockerin Type I: 52 kDa

Nature Methods: doi:10.1038/nmeth.3099

Supplementary Table 3. Yield of interpretable curves

Construct	Interpretable Curves
GFP	25 out of 15258 = 0.16 %
Fibronectin	27 out of 26653 = 0.1 %
Xylanase	91 out of 5553 = 1.64 %
Spectrin	50 out of 10344 = 0.48%

References

1. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).
2. Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M. & Scheuring, S. High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science* **342,** 741–743 (2013).

Nature Methods: doi:10.1038/nmeth.3099

Chapter 6

Elastin-Like Polypeptides in Single Molecule Force Spectroscopy

6.1 Summary

The discovery and characterization of high-force complexes in single molecule force spectroscopy (SMFS) opened a new force-regime for experiments. In site directed immobilization approaches polyethylene glycol (PEG) is the standard linker polymer. PEG serves two purposes. First, it acts as spacer between the target protein and the surface to passivate against any interfering protein:surface interactions. Second, commercially available PEG has a wide variety of reactive groups that can be used in protein pull-down techniques.
However, PEG undergoes a trans-trans-gauche to all-trans conformation under force [97]. At low forces water molecules stabilize PEG in a more compact form (trans-trans-gauche) *via* hydrogen bonds. Whenever force is applied to the fused protein complex the PEG polymer is loaded as well, increasing the probability of the conformational change of PEG. The higher the force, the higher the probability of overall PEG conformational change which leads to an increase in the net contour length recorded in an experiment. Hence, the conformational change of PEG biases the real contour length of the particular stretched domains. To solve this issue more inert linker molecules are needed.
Associated publication P7 describes a new PCR-based approach to clone and purify these biological linkers. Elastin-like poylpeptides (ELPs) meet the requirements for linkers in single molecule force spectroscopy (SMFS) studies: It is possible to create ELPs with defined length and monodisperse composition for custom made linker molecules in SMFS experiments. Additionally they are easy to produce and to purify [134].
Basis for the approach is a short, synthesized ELP gene, which is amplified by PCR and ligated *via* GoldenGate Assembly [110]. Up to three fragments can be joined in one assembly reaction, which enables a screening-free modular assembly of ELPs with defined length.
Associated publication P8 employs the customized ELPs as linkers in an atomic force microscope (AFM) experiments. ELPs contain a cysteine and a Sortase A-recognition sequence at opposite ends. Hence, they can be immobilized on a maleimide activated surface *via* their cysteine and the proteins of interest are enzymatically fused to the ELPs by Sortase A. Thus all components that are loaded with force are fully protein-based. Compared to PEG, ELPs showed advantageous behavior, as SMFS experiments yielded homogeneous, monodisperse and accurate force-distance traces across all force regimes probed.

6.2 Associated Publication P7

Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation

by

Wolfgang Ott, Thomas Nicolaus, Hermann E. Gaub, and Michael A. Nash

published in

Biomacromolecules, 17 (4), 1330-1338, (2016)

Article
pubs.acs.org/Biomac

Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation

Wolfgang Ott,[†,‡,§] Thomas Nicolaus,[†] Hermann E. Gaub,[†,‡] and Michael A. Nash*[,†,‡,∥,⊥]

[†]Lehrstuhl für Angewandte Physik and [‡]Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität München, 80799 Munich, Germany

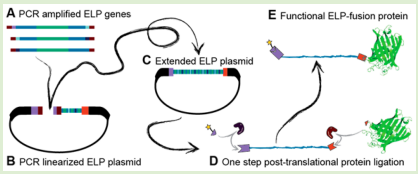
[§]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany

[∥]Department of Chemistry, University of Basel, 4056 Basel, Switzerland

[⊥]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH-Zürich), 4058 Basel, Switzerland

Supporting Information

ABSTRACT: Repetitive protein-based polymers are important for many applications in biotechnology and biomaterials development. Here we describe the sequential additive ligation of highly repetitive DNA sequences, their assembly into genes encoding protein−polymers with precisely tunable lengths and compositions, and their end-specific post-translational modification with organic dyes and fluorescent protein domains. Our new Golden Gate-based cloning approach relies on incorporation of only type IIS *BsaI* restriction enzyme recognition sites using PCR, which allowed us to install ybbR-peptide tags, Sortase c-tags, and cysteine residues onto either end of the repetitive gene polymers without leaving residual cloning scars. The assembled genes were expressed in *Escherichia coli* and purified using inverse transition cycling (ITC). Characterization by cloud point spectrophotometry, and denaturing polyacrylamide gel electrophoresis with fluorescence detection confirmed successful phosphopantetheinyl transferase (Sfp)-mediated post-translational N-terminal labeling of the protein−polymers with a coenzyme A-647 dye (CoA-647) and simultaneous sortase-mediated C-terminal labeling with a GFP domain containing an N-terminal GG-motif in a one-pot reaction. In a further demonstration, we installed an N-terminal cysteine residue into an elastin-like polypeptide (ELP) that was subsequently conjugated to a single chain poly(ethylene glycol)-maleimide (PEG-maleimide) synthetic polymer, noticeably shifting the ELP cloud point. The ability to straightforwardly assemble repetitive DNA sequences encoding ELPs of precisely tunable length and to post-translationally modify them specifically at the N- and C- termini provides a versatile platform for the design and production of multifunctional smart protein−polymeric materials.



INTRODUCTION

Repetitive polymers of controlled length and tunable phase-transition behavior are urgently needed for a variety of applications in the nano/biosciences, including drug delivery[1,2] and medical diagnostics.[3] Such stimuli-responsive polymeric materials are of high interest for fundamental investigations into biomolecules under the influence of mechanical, thermal, and chemical denaturants using biophysical methods such as single-molecule AFM force spectroscopy[4,5] and microscale thermophoresis.[6] Elastin-like polypeptides (ELPs) are artificial proteins derived from naturally occurring elastomeric proteins.[7,8] Recombinant ELPs consist of repeats of the amino acid sequence Val-Pro-Gly-Xaa-Gly, where Xaa represents any amino acid except proline. ELPs exhibit a reversible lower critical solution temperature (LCST) and undergo a phase transition that can be triggered by temperature.[9] Other environmental stimuli like pH or ionic strength can also be used to collapse ELPs under isothermal conditions. The guest residue (Xaa) can be used to influence the pH/thermal phase transition properties of the resulting protein−polymers. Insertion of acidic residues such as glutamate or aspartate in the guest residue position results in pH-responsive behavior. The transition temperature is strongly dependent on the concentration and molecular weight, with longer ELP sequences collapsing at lower temperatures. One can also tune the cloud point by changing several environmental parameters at once (e.g., temperature, pH, salt), thereby shifting the transition to lower or higher temperatures as desired.[10]

These unique properties of ELPs make them attractive for a variety of applications and scientific investigations.[11] Chromatography free protein purification, for example, can be

Received: December 21, 2015
Revised: February 25, 2016
Published: March 14, 2016

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

Table 1. Overview of Employed Primers[a]

primer	sequence 5′−3′
(1a) FW ELP I ybbR	TATATAGGTCTC*CTGGC***TGTGCCGGGA**GAAGGAGTCCCTGG**TGTCGGTGTCCCAGGCG**
(1b) REV ELP I	GGTCTC*CTCCTT***CACCCGGAACGCCACCCCCCGGAACACCGCCGC**
(2a) FW ELP II	TATATAGGTCTC*AAGGA***GTACCAGGCGAAGGCGTGCCGGGTGTC**
(2b) REV ELP II	ATATATGGTCT**C***ACCCT***CACCCGGAACG**CCC**ACCCCCCGGAACACCGCCGC**
(3a) FW ELP III	TATATAGGTCTC*GAGGGT***GTACCAGGCGAAGGGGTGCCGGGTGTC**
(3b) REV ELP III LPETGG	ATATATGGTCTC*CGGCA*GACCTT**CACCCGGAACGCCACCCCCCGGAACACCGCCGC**
(4) REV ELP III	ATATATGGTCTC**CACCTT**CACCCGGAACGCCACCCCCCGGAACACCGCCGC
(5) FW backbone LPETGG	ATATATGGTCTC*CTGCCG*GAAACCGGCGGC**TAACTCGAGTAAGATCCGGCTGC**
(6) REV backbone ybbR	ATATATGGTCTC**AGCCAGTTTAGAAGCGATGAATTCCAG**
(7) FW backbone ybbR	**GACTCTCTGGAATTCATCGCTTCTAAACTGGCT**GGTCTC*CAGGT***GTGCCGGGA**
(8) FW ELP II ybbR	TATATAGGTCTC*CTGGCG***GTACCAGGCGAAGGGGTGCCGGGTGTC**
(9) FW ELP III ybbR	TATATAGGTCTC*CTGGCG***GTACCAGGCGAAGGCGTGCCGGGTGTC**
(10) FW ELP N Cys	**GACTCTCTGGAATTCATCGCTTCTAAACTGGCT**GGTCTC*CTGCG***TGCCGGGAGAAGGAG**
(11) REV backbone	**CCCGGCACAGCCAGTTTAGAAGCGATGAATTCCAGAGAGTC**GGTCTC*ACATATG***TATATC**

[a]Primers 1−7 are employed for the cloning of the ELPs with three fragments growth every cycle. Primers 1−4 are necessary for insert amplification and primers 5−7 are used for amplification of the backbone. Primers 8 and 9 are only important for ELP cloning procedures with the addition of one or two fragments. Primers 10 and 11 were used to change the 5′ flanking site of the ELP gene from the gene for the ybbR-tag to a cysteine. DNA sequence is styled in different ways: **bold** (annealing region), underlined (*Bsa*I recognition site), and *italic* (*Bsa*I restriction site).

performed by producing a target protein as an ELP fusion and precipitating it from cellular extracts, avoiding the need for affinity tags. This method allows for purification of recombinant proteins under mild conditions. Moreover, it is reported that, in combination with maltose binding proteins, ELPs can improve the solubility of fusion domains and thereby improve expression yields.[12−14]

In the field of biomaterials science, ELPs represent a viable option as a scaffold material for tissue engineering or as carriers for drug molecules. Applications for in vivo systems demand high predictability and controllability of the biophysical behavior of the molecules. Since ELPs consist only of amino acids, they are competitive in terms of biocompatibility and biodegradation in vivo as compared to their synthetic organic polymer counterparts.[15,16] ELPs possess the added advantage of complete monodispersity. More fundamentally, the phase transition characteristics of ELPs have served as an ideal model system for theoretical calculations and modeling studies.[17−21] Additionally, conjugates of ELPs and synthetic polymers (e.g., PEG) are of high interest and benefit from site-specific conjugation approaches.[22,23]

In order to fully leverage the versatility of repetitive protein−polymers such as ELPs, modular and straightforward approaches to cloning and site-specific post-translational modification are highly desirable. Standard solid-phase gene synthesis methods are, so far, not able to produce the long (>600 bp) strands of repetitive DNA required for encoding thermally responsive elastin-like polypeptides (ELPs) with lengths >200 amino acids. Typically rationally designed ELPs are constructed using recursive directional ligation (RDL), which requires plasmid amplification and restriction digestion and imposes certain restrictions (i.e., the absence of restriction sites).[24] Larger ELP genes can also be obtained with the OERCA (overlap extension rolling circle amplification) method, which generates a distribution of unspecified lengths of repetitive DNA sequences.[25]

Compared to the RDL method, our Golden Gate approach presented here avoids cloning scars due to the use of type IIS restriction enzymes and is able to cut scarlessly within the coding region.[24,26] The PRe-RDL (RDL by plasmid reconstruction) method relies on several type IIS restriction enzymes and requires certain modifications of the backbone beforehand.[27]

Our method is applicable to a broad spectrum of plasmids, since the only limitation is one type IIS restriction enzyme with a recognition site not present in the backbone. Along with this advantage, it is likewise ideal for adding ELPs to an existing gene-containing plasmid to create fusion proteins with different length ELPs. The combinational possibilities also do not rely on a plasmid library, but can be designed using a bottom-up block assembly approach. Our approach can also be used in a complementary way with the existing RDL and OERCA methods, for example, by easily generating fast and reliable plasmid libraries which can then be further extended by combining with RDL or OERCA methods.

We present a sequence independent approach based on the Golden Gate technology employing polymerase chain reaction (PCR) amplification of short ELP repeats and ligation into a plasmid backbone to produce repetitive ELP genes with specific peptide tag end groups for covalent post-translational modification. A single type IIS restriction enzyme is used to create unique ends and guarantee the order of DNA block assembly. Using this method, repetitive DNA sequences up to hundreds of nm in length (i.e., 120 pentapeptide repeats of ELPs) can be rationally designed and created. The 5′ and 3′ peptide tags for post-translational modifications were readily incorporated during the cloning workflow, providing many further possibilities for downstream conjugation and labeling. We were able to install a ybbR[28] tag and sortase c-tag to the ELP, enabling enzyme-catalyzed ligation to fluorescent proteins and organic dyes (as shown below). Our approach builds on the prior method shown by Huber et al. which demonstrated fusion of different kinds of repetitive DNA to create chimeras of ELPs, silk peptides, and similar proteins.[29] Our methodology is also compatible with their approach with the advantage of using only one type IIS restriction enzyme.

Alternatively, it is possible to modify the carrier plasmid in the first amplification round and add ELP flanking tags or protein domains easily. Since the reaction starts new every three fragments, one can easily define block patterns that build up an overall sequence. For example, pH responsive blocks can be interspersed with pH-insensitive blocks. In regard to user-

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

friendliness, the presented method is advantageous because it relies on the same ELP gene inserts, which can be reused. Once successful amplification and purification of the sequences is achieved, the PCR amplicons can be stored and used again as needed. This way it is possible to create a whole library of gene sequences and, if desired, shuffle these each ligation cycle. Post-translational fusion of ELPs using Sortase ligation circumvents the known issue of low protein yields for N-terminally located ELP domains in fusion proteins.[30,31] Instead of optimizing expression conditions for proteins of low yield, a protein of interest can be produced in its native state and fused afterward post-translationally with the ELP domain. To the best of our knowledge, this represents the first report using a Sortase-based recognition sequence to fuse ELP proteins to other proteins.[13,32]

MATERIALS AND METHODS

All used reagents were of analytical purity grade and were purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.) or Carl Roth GmbH (Karlsruhe, Germany).

Monomer Gene Synthesis. A synthetic gene encoding 150 nucleotides (10 pentapeptide repeats) for the $(VPGVG)_5$-$(VPGAG)_2$-$(VPGGG)_3$ peptide (Centic Biotech, Heidelberg, Germany) served as starting material (see Supporting Information, DNA Sequence 1 and Protein Sequence 1).

Cloning. Golden Gate cloning was employed to create the different rationally designed ELP constructs.[26] PCR (Backbone: 98 °C 2 min, (98 °C 7 s, 72 °C 2 min 30 s) x30, 72 °C 5 min; Insert: 98 °C 2 min, 98 °C 7 s, 60 °C 7s, 72 °C 5 s) x30, 72 °C 5 min) was performed with a Phusion high fidelity polymerase master mix. A typical 20 μL PCR mix contained 10 μL of Phusion high fidelity polymerase master mix (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.), 0.5 μL per forward and reverse primer (10 μM), 1.5 μL DMSO, 1 ng of template, and water. All primers (biomers.net, Ulm, Germany) used in this study are listed in Table 1.

In the first round of PCR (see backbone PCR above, 55 °C 7s annealing), the backbone of a modified pET28a vector (Merck KGaA, Darmstadt, Germany) was linearized. The PCR product contained at the 5′ end the sequence for a ybbR-tag (DSLEFIASKLA) and at the 3′ end a C-terminal Sortase recognition sequence (LPETGG).[33,34] Sequences of all PCR fragments (backbone, ELP I, II, III, IV) and a description for primer design (see Supporting Information, Primer 12) based on an original pET28a vector are attached in the Supporting Information (Figures S1−S9, DNA sequences 1−6 and Figures S14−S18).

The superfolder GFP (sfGFP) plasmid was created with Gibson Assembly.[35] The gene (Addgene ID: 58708)[36] was amplified with overlaps to match a linearized vector containing sequences encoding N-terminal HIS_6-tag, a TEV protease cleavage site, and two glycines (compare the PCR program above; 55 °C annealing and an extension time of 1 min 30 s; see Supporting Information, DNA sequence 8 and protein sequence 4).

All PCR products were digested (37 °C, 1−12 h) with FD-*Dpn*I (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.) and purified either with QIAquick PCR purification kit or gel extraction kit (Qiagen, Hilden, Germany; Supporting Information, Figures S10 and S13). *Dpn*I was added to digest the methylated plasmids serving as starting material (template) in the PCRs, to reduce number of false positive clones in the following transformation.

Typically, a 25 μL Golden Gate reaction (2.5 μL CutSmart buffer (10×), 1.25 μL T7 ligase, 1.25 μL *Bsa*I-HF and 2.5 μL ATP (10 mM), New England Biolabs, Ipswich, MA, U.S.A.) was set up. The inserts were added in 10-fold molar excess to the backbone (ratio of 0.1 pmol insert to 0.01 pmol backbone). The reaction was performed in a thermo cycler (25× 37 °C 2 min, 25 °C 5 min; 37 °C 10 min; 80 °C 10 min). For the Gibson Assembly reaction, 10 μL of the master mix (2×, New England Biolabs, Ipswich, MA, U.S.A.) were mixed with 0.01 pmol vector and 0.1 pmol insert. The reaction was incubated for 1 h at 50 °C. For the replacement of the ybbR-tag with cysteine, the PCR linearized product was first digested with *Bsa*I-HF together with FD-*Dpn*I (1 h, 37 °C, 5 min, 80 °C). The reaction was supplied with 1 μL of dNTPs (10 mM, New England Biolabs, Ipswich, MA, U.S.A.), 1 μL of Klenow Fragment (10 U/μL, Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.), and incubated (37 °C, 15 min, and 75 °C, 10 min). After a gel extraction, the corresponding band was excised and a blunt end reaction (6.5 μL PCR product, 1 μL ATP (10 mM), 1 μL CutSmart buffer (10×), 0.5 μL PEG-6000, 1.0 μL T4 Polynucleotide Kinase, 1.0 μL T4 Ligase) was set up (37 °C 15 min, 22 °C 45 min, 80 °C 7 min).

In case of the Golden Gate reaction, 10 μL, and in case of the Gibson Assembly or the blunt end ligation, 2 μL, were used to transform DH5α cells (Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 42 °C 1 min, 1 h 37 °C). The transformed culture was plated on appropriate antibiotic LB-Agar plates. A small number (<10) of clones were analyzed by colony PCR, or analytical restriction digestion (FD-*Eco*RI, Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.) followed by sequencing (Supporting Information, Table S1).

Protein Expression. For ELP expression, chemically competent *E. coli* NiCo21(DE3) (New England Biolabs, Ipswich, MA, U.S.A.) were transformed with 50 ng plasmid DNA.[37] The cells were incubated in kanamycin containing, autoinducing ZYM-5052 media (supplemented with an amino acid mix 0.1 mg/mL) 24 h at 25 °C.[38−40] After harvesting, ice cooled cells were lysed using sonication (Bandelin Sonoplus GM 70, Tip: Bandelin Sonoplus MS 73, Berlin, Germany; 40% power, 30% cycle 2 × 10 min). The supernatant of the lysate (15000 g, 4 °C, 1 h) was heated to 60 °C for 30 min to denature most of the *E. coli* host proteins. In a second step, the collapsed ELPs within this clouded solution were rehydrated by incubating under continuous mixing for 2 h at 4 °C. This allowed the resolubilization of the ELPs while the precipitated host proteins remained insoluble. A centrifugation step (15000 g, 4 °C, 30 min) was used to separate the soluble ELPs and remaining proteins from precipitated cell debris. The clear supernatant turned immediately cloudy after adding 1 M acetate buffer (final concentration 50 mM, pH 3.5), and 2 M NaCl in crystalline form. The mixture was incubated for 30 min at 60 °C. The collapsed ELPs were collected by centrifugation (3220 g, 40 °C, 75 min). The obtained pellet was resolubilized in 50 mM Tris-HCl (pH 7.0) and incubated overnight at 4 °C. The remaining precipitated debris were removed by a final centrifugation step (3220 g, 4 °C, 60 min). The supernatant was mixed again with acetate buffer and sodium chloride to collapse the ELPs. After the heated incubation and centrifugation step, the pellet was resolubilized in buffer (50 mM Tris-HCl, pH 7.0).[14,41]

The purity of the ELP was confirmed by SDS-PAGE (Any kD Mini-PROTEAN Stain-Free Gels, Bio-Rad Laboratories GmbH, Hercules, CA, U.S.A.), in order to detect any remaining contaminant host proteins. The ELPs were labeled with CoA-647 (New England Biolabs, Ipswich, MA, U.S.A.) and Sfp (37 °C, 1 h, 5 mM $MgSO_4$) to visualize them. After labeling, the ELPs were mixed with 6× loading buffer and heated to 95 °C for 10 min.[42] Usually a purity grade of >95% was obtained. Purity analysis was performed by overlaying the UV active Stain-Free technology from Bio-Rad (labeling all tryptophan side groups of *E. coli* host proteins) and a fluorophore specific red channel for the CoA-647-ELP constructs (Supporting Information, Figure S11). MALDI-TOF analysis of ELP samples ELP_{30-50} was performed to increase confidence in the high purity of the samples (Supporting Information, Figure S19). ELPs were stored at 4 °C in 50 mM Tris-HCl, pH 7.0.

The final ELP concentration was photometrically determined at 205 nm (Ultrospec 3100 pro, Amersham Biosciences (Amersham, England) and TrayCell (Hellma GmbH & Co. KG, Müllheim, Germany)).[43]

For the expression of HIS_6-TEV-GG-sfGFP, 50 ng plasmid DNA was used to transform *E. coli* NiCo21(DE3) cells. Kanamycin containing, autoinducing ZYM-5052 growth media was inoculated with an overnight culture.[38] After 24 h incubation at 25 °C, the cells were harvested, lysed, and centrifuged as described above. The

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

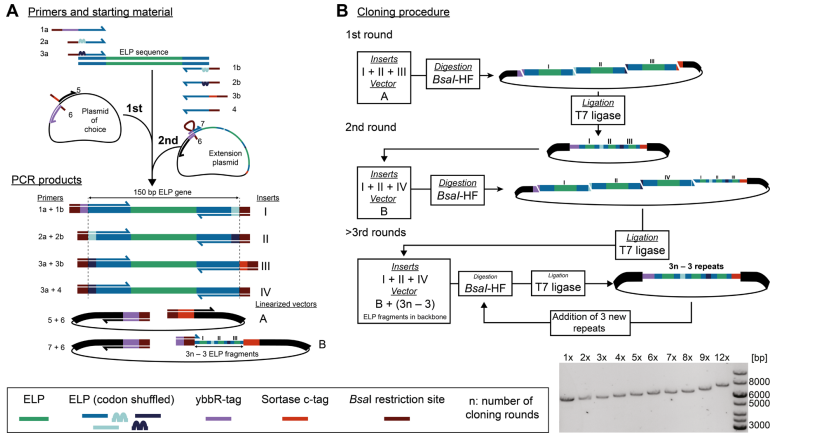


Figure 1. Cloning schematic: (A) The schematic describes the process of sequence independent PCR amplification of unique inserts (I–III) from the same template. The amplification of the first backbone (plasmid A) enables subcloning of the first three inserts, which leads to plasmid B. Plasmid B is linearized at the N-terminal ybbR-tag, as are all the following backbones. The new ELP amplicons can always be inserted upstream of the old ELP repeats. (B) Repetitive rounds of cloning add subsequently more ELP inserts until the desired length is achieved.

supernatant was applied on a HisTrap FF (GE Healthcare Europe GmbH, Freiburg, Germany). After washing five times with wash buffer (25 mM Tris-HCl pH 7.8, 300 mM NaCl, 20 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol), the bound protein was eluted (25 mM Tris-HCl pH 7.8, 300 mM NaCl, 300 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol).

HIS_6-TEV-GG-sfGFP fusion protein (TEV cleavage site: ENLYFQG) was dialyzed immediately after elution with the TEV protease (4 °C, 50 mM Tris-HCl, pH 7.0) overnight. The cleaved product was separated from the uncleaved construct by applying the reaction mix on a HisTrap FF 5 mL column. The successfully cut fragment in the flow through was collected. The fraction was dialyzed against 50 mM Tris-HCl, pH 7.0, and stored in 50% (v/v) glycerol at −80 °C. The purity of the elution and the cleaved fraction was analyzed via a SDS-PAGE analysis. The specific extinction coefficient of GFP at 485 nm was used to determine the concentration of GG-sfGFP.

Turbidity Measurements. For the turbidity measurements, a photometer with a Peltier heating element was used (JASCO V-650, JASCO Germany GmbH, Gross-Umstadt, Germany). The turbidity was determined at 350 nm, while the temperature was ramped at a rate of 2 °C/min. Measurements were taken every 0.5 °C between 20 and 80 °C. ELPs were dialyzed against double distilled water, diluted into 50 mM Tris-HCl, pH 7.0, followed by addition of sodium chloride to achieve the desired final concentration.

For NaCl titration, 100 μM of the ELP constructs were tested in a range of 0–3 M sodium chloride. The 6× ELP construct was also probed in a concentration range of 25–200 μM with different NaCl concentrations.

For pH titrations, stock solutions of 0.1 M phosphate-citrate buffer at different pH values were mixed with solutions of water solubilized ELPs. Hereby a final concentration of 0.05 M of the phosphate-citrate buffer was obtained.

Data analysis of the transition temperature curves (for NaCl, pH, concentration dependency, and PEG-ELP fusions) was performed by fitting the measured data points with a four-parameter logistic function to obtain the corresponding transition temperature.

Sortase and Sfp-Mediated Protein Ligation. For highest ligation efficiencies, enhanced Sortase (eSortase) was used in the reaction.[44] The reaction conditions for both Sfp and eSortase enzymes were chosen according their reported reaction maxima to achieve highest activities.[28] ELPs in excess were added to a solution containing 50 mM Tris-HCl, pH 7.5, 15 μM ELP, 0.5 μM GG-sfGFP, 0.2 μM eSortase, 1 μM Sfp, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 5 μM CoA-647. The ligation reaction was incubated for 2 h at 37 °C.

Cysteine-Maleimide Bioconjugation Reaction. Cysteine-containing ELPs were reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP, (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.)). After the removal of TCEP with Zeba Spin Desalting Columns 7K (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.) cysteine-ELPs were mixed with $Alexa_{647}$-C2-Maleimide (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.) and incubated for 1 h at 37 °C (100 mM Tris-HCl, pH 7.0; Supporting Information, Figure S20).

PEG (MW: 20000 Da, α-methoxy-ω-maleimide, Rapp Polymere GmbH, Tübingen, Germany) was used in different molar ratios in the bioconjugation reaction with cysteine-ELP_{60} or ELP_{60}. A total of 75 μM of the reduced ELPs were mixed with Tris-HCl (pH 7.0, 100 mM), PEG, and incubated for 1 h at room temperature. After that they were mixed with 5 M NaCl and to a final concentration of 3 M NaCl, and their cloud point was determined as described above.

RESULTS AND DISCUSSION

Our sequence-independent Golden Gate-based method provides an easy way to create defined repetitive DNA sequences.[26] We designed and produced gene cassettes encoding repetitive proteins several hundreds of amino acids in length. Figure 1 outlines the principle of primer design and the following logical and stepwise workflow. The sequence of the starting synthetic gene was designed in such a way that the codon usage within the first and last 15 nucleotides was unique within the otherwise repetitive 150 bp sequence. This was necessary to ensure specific annealing of primers at the 5′ and 3′ end. Desired modifications were introduced by overhangs of the primers at their 5′ end (i.e., *BsaI* recognition site) or at their

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330–1338

Table 2. Biophysical Properties of the Characterized ELP Constructs

ELP repeats $(5)_x$	ε_{205}[43] (1/M cm)	mol wt[45] (Da)	glutamate residues in ELP repeat	isoelectric point[45]	amino acids in ELP repeats (total)	total length[46] (nm)
10	196690	5893.7	2	3.91	50 (68)	24.82
20	335690	9908.2	4	3.77	100 (118)	43.07
30	474690	13922.8	6	3.67	150 (168)	61.32
40	613690	17937.3	8	3.59	200 (218)	79.57
50	752690	21951.9	10	3.53	250 (268)	97.82
60	891960	25966.4	12	3.47	300 (318)	116.07
Cys-60	855980	24894.2	12	3.20	300 (308)	112.42

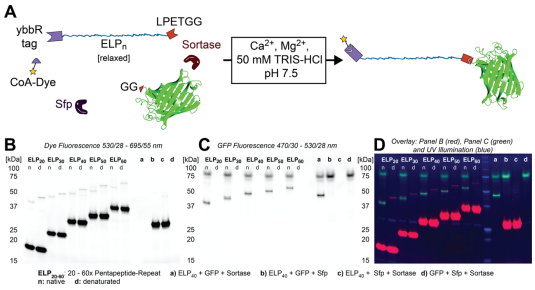


Figure 2. Post-translational ligation of the ELP peptide. (A) Schematic of the ELP constructs containing a N-terminal ybbR-tag and a C-terminal Sortase-tag. A post-translational one-pot reaction was used to fuse a CoA-647 fluorescent dye to the N-Termini via an Sfp-catalyzed reaction. In parallel, the eSortase fuses a GG-sfGFP toward the C-terminal LPETGG. (B) An image of a SDS gel obtained following dual labeling of ELPs under different reaction conditions and ELP lengths. The image shows only the red CoA-647 dye (ex: 530/28, em: 695/55 nm). (C) Fluorescent image of the same gel as in B, but this time with blue excitation (ex: 470/30, em: 530/28 nm), hence, only the native GFP specific bands are visible. (D) Overlay of B and C plus additional UV illumination which excites tryptophan side group converted fluorophores enabled by the Bio-Rad Stain-Free technology.

3′ region (codon shuffling of nucleotides). It was then possible to create 150 bp ELP genes with different flanking regions from the same template (Primers 1−3; Figure 1A) using PCR primers that annealed at the 5′ and 3′ ends of the synthetic gene.

In the first amplification and linearization reaction of the plasmid, primers annealed at the desired ELP gene insertion site, that is, at the opening location on the plasmid during the first PCR. In our case this was downstream of the T7 promoter and upstream of the T7 terminator (see Supporting Information, Figures S8 and S9). However, due to the freedom of primer design and plasmid choice, the insertion site can in principle be anywhere in the plasmid. The primers linearized the plasmid and introduced tags at the 5′ (ybbR-tag) and 3′ (Sortase c-tag) ends, as well as *BsaI* recognition sites (Primers 5 and 6; vector A; Figure 1 A). In our case, a modified pET28a vector, already containing a ybbR-site downstream of the T7 promotor immediately following the start codon AUG, served as template. Hence, only the Sortase c-tag was newly introduced (see Supporting Information for primers for the standard pET28a vector). The continuing general ELP expansion principle relies on having three different PCR amplified ELP fragments (I−III) with different codon usages at their 5′ and 3′ end, within the *BsaI*-restriction site (Supporting Information, Figures S1−S6). This design made logical and block-wise gene assembly possible. The selected primers introduced a shuffled 3′ end that matched the 5′ end of the subsequent fragment. In the first ELP assembly round, the 5′ end of fragment I matched the ybbR-tag of the linearized backbone. The 3′ end of fragment III had compatible sticky ends with the Sortase c-tag of the linearized plasmid (Figure 1B, first round). After successful annealing of sticky ends, the T7 ligase covalently linked the three ELP fragments seamlessly into the plasmid without any undesirable cloning scars in between.

The forward primer (Figure 1A, second: primer 7) for the following plasmid linearization rounds annealed at a different site within the ELP-containing plasmid, compared to the initial linearization round (Figure 1A, first: primer 5). It annealed at the ybbR-tag and the 5′ end of the ELP gene. Right in between the two coding regions, a nonannealing loop encoding a *BsaI* recognition site was introduced (Figure 1A, second, and Figure 1B, second round) with the primer. The annealing at the ybbR-tag was necessary to ensure high temperature-dependent primer annealing specificity at the very 5′ end of the ELP gene; otherwise, the primer would anneal at every fragment I throughout the whole assembled ELP gene cassette. High annealing temperatures minimize undesired PCR side products, that is, only partly ELP-containing, linearized vectors. The reverse primer was the same for all plasmid linearization reactions (Figure 1A; vector B). After the restriction digestion reaction, the linear plasmid now had a Sortase c-tag sticky end at the 3′ end and an ELP fragment I sticky end at the 5′ end.

Now only the last ELP fragment (Figure 1A, insert IV) had to be amplified with a different reverse primer (Figure 1A, primer 4) to yield a PCR product with a compatible 3′ end to

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

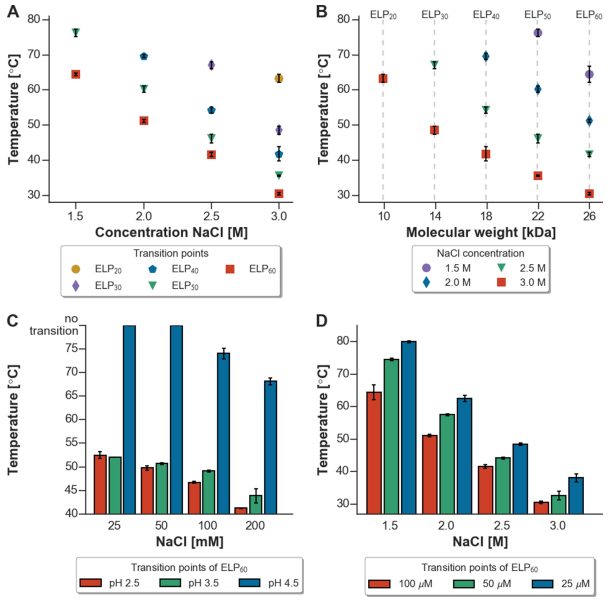


Figure 3. Cloud point characterization of the 10−60 pentapeptide ELP repeats. (A) Characteristic decrease in the transition temperature of the ELPs with increasing sodium chloride concentration and ELP length. (B) Relation between decreasing transition temperature and increasing molecular weight. (C) Correlation between pH, NaCl, and transition temperature. (D) Concentration dependency of the transition temperature for the 60 pentapeptide ELP repeat. Data points for the plots were obtained from triplicates. Error bars represent the standard deviation of the average transition point.

the already existing ELP cassette. The growing ELP insert in the plasmid always started with fragment I. This made the reuse of the amplified insert sequences (I, II, IV) for every following expansion cloning round possible (Figure.1B, >third rounds).

This method not only allows a logical assembly of repetitive gene patterns, but also makes the modification of flanking regions or mutation of the first base pairs at 5′ end 3′ end possible. For example, we introduced two glutamates in each of the fragments at their 5′ and 3′ ends by changing the codon from the "X" guest residue at the 5′ and 3′ end of the VPGXG motif to a glutamate (VPGEG). The primers did not align completely with the template and introduced the glutamate mutation during PCR amplification. The chemically synthesized sequence also had some minor mistakes at the 3′ end, which were corrected with primers within the initial PCR. The final ELP substructure of all ELPs used in this study consisted of 10 pentapeptide repeats ($VPGXG_{10}$, X being $[EV_4A_2G_2E]$). For the rest of the manuscript this motif is referred as ELP_n, with n being the number of pentapeptide repeats of this motif (see Supporting Information, DNA sequence 2, protein sequence 2, and DNA sequence 7, protein sequence 3).

We ligated three 150 bp fragments with a linearized vector of choice in one step. It was possible to modify the 5′ and 3′ ends of the fragments with overhang primers prior to ligation, in our case with an N-terminal ybbR and a C-terminal Sortase tag (Figure 1B). Overall, seven different ELP constructs were used in this study for biophysical characterization of the peptide sequence, while ten were successfully cloned. The largest ELP gene contained 120 pentapeptide repeats. All ELP constructs were built with the four different ELP PCR products from the same batch. PCR gels from the fragments and an overview of cloning efficiencies can be found in the Supporting Information (Figure S9 and Table S1). Typical yields after the purification were 56−138 mg protein/l culture, while the ELP_{10} repeat had the lowest yield (2 mg protein/l culture).

Table 2 shows biophysical characteristics of the ELPs characterized in this study. Each ELP was produced with a ybbR-tag at the N-terminus and a Sortase c-tag at the C-terminus. In the bottom right corner of the schematic (Figure 1), FD-*Eco*RI digested plasmids are shown on an agarose gel. The gel analysis shows the successful construction of plasmids containing 10 to 120 pentapeptide repeats.

Following successful cloning, expression and purification, we tested the functionality of the attached terminal tags. Figure 2A shows the scheme for post-translational protein ligation reactions. The ELPs of varying lengths contain an N-terminal ybbR-tag and a C-terminal Sortase recognition sequence (i.e., LPETGG). Figure 2B and C show an SDS-PAGE image of the same gel with different excitation and emission filters. Using a reaction catalyzed by Sfp, it was possible to fuse a fluorescently

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

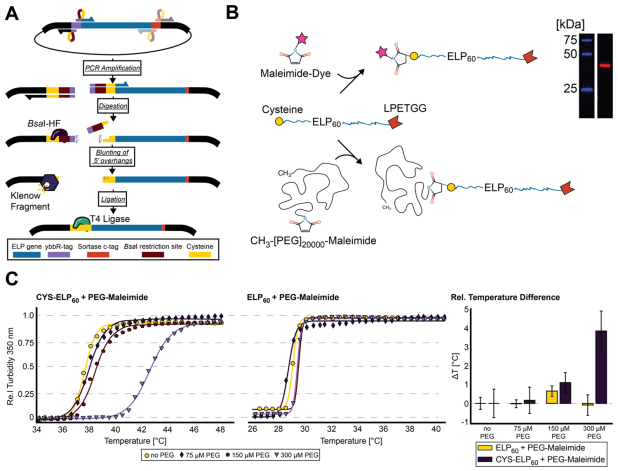


Figure 4. Cloning schematic and bioconjugation of cysteine-ELPs with a maleimide-dye and PEG-maleimide. (A) Illustration of the cloning schematic for changing the ELP flanking regions (i.e., replacement of the ybbR-sequence with a cysteine). The ybbR-sequence was deleted via a PCR reaction, and a cysteine was introduced (see primers 10 and 11 in Table 1). The flanking restriction sites were digested with *Bsa*I-HF and the remaining sticky ends were filled in with Klenow fragment. Finally, the linear product was circularized with T4 ligase. (B) Procedure of the bioconjugation reaction. Cysteine-ELPs were reduced with TCEP and conjugated to a maleimide dye or a PEG-maleimide polymer. A gel image on the right shows the successful conjugation reaction between the dye and the ELP. (C) Bioconjugation of a 20 kDa PEG-maleimide to CYS-ELP_{60} shifted the cloud point up by ~4 °C (left panel). The cloud point of ELP_{60} lacking cysteine (middle panel) was not influenced by the addition of maleimide PEG. The right panel shows the cloud point shift (ΔT) due to addition of different concentrations of PEG-maleimide. Data points for the plot were obtained from triplicates. Error bars account for Gaussian error propagation due to calculation of the difference of the average transition point from three samples.

labeled CoA-647 to the ELP (N-terminal ybbR-tag). Results of the specific excitation for the CoA-647 dye are shown in Figure 2B. Brightest are the CoA-647-ELP fusions proteins, but also the CoA-647-ELP-sfGFP fusion proteins are visible above the bright monomer band. Fully denatured proteins appear slightly higher in the gel due to their different running behavior. The Sortase-tag was simultaneously utilized for fusion of different proteins to the ELP sequences (C-terminal LPETGG). A GG-sfGFP was fused to the ELPs, which was excited with blue LED light and detected within the green emission of sfGFP (Figure 2C). Nonligated and nondenaturated GFP appears at the top of the gel, since it does not run according its molecular weight in its native (i.e., correctly folded) state (see Supporting Information, Figure S12). No GFP fluorescence is visible in the heated samples due to complete denaturation of the GFP. Figure 2D shows an overlay of Figure 2B and C, visualizing the successful post-translational ligation of GG-sfGFP and CoA-647 to the different ELP peptides within a one-pot reaction. The ligation efficiency of the Sortase never goes to 100% completion. Due to the Sortase reaction mechanism, a dynamic equilibrium is eventually reached and complete fusion of GG-sfGFP to ELP is therefore not to be expected.[47]

After confirming the biochemical accessibility and functionality of the terminal ybbR- and Sortase-tags, we characterized the phase behavior of the modified ELPs. Figure 3 presents an overview of the lower critical solution temperatures (LCSTs) of the characterized ELPs under various conditions. First the temperature dependence of the $ELP_{10\text{-}60}$ constructs were probed against different sodium chloride concentrations, at neutral pH (50 mM Tris-HCl, pH 7.0; Figure 3A). The 10 pentapeptide repeat ELP did not collapse below 80 °C, which is in agreement with the remainder of the data set if one looks at the increasing transition temperature with decreasing size of the construct. The 20 pentapeptide ELP repeat, for example, only collapsed with 3 M of sodium chloride at 60 °C. Figure 3B clarifies the correlation between salt concentration, molecular mass and transition temperatures. Only the longest ELP construct collapsed across all given sodium chloride concentrations in the temperature range from 20 to 80 °C. Salt-induced cloud point shifts are a well-known characteristic of ELPs.[15,24,48]

The incorporation of two glutamates per ten pentapeptides resulted in pH-dependent transitions. ELPs with glutamates were expected to show pH-responsiveness. Above their pK_a the ELPs have a relatively high transition temperature, since the glutamates are deprotonated and ionized and therefore electrostatically repel each other. Below or close to their corresponding pK_a, the transition temperature significantly decreases due to protonation and neutralization of the negative charge (Figure 3C). The decreasing influence of salt at lower pH is similar to that demonstrated by MacKay et al.[49] Figure 3D illustrates the dependence of transition temperature on the ELP concentration. At concentrations above 100 μM, the 60 pentapeptide ELP (150 and 200 μM) already collapsed at room

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

temperature; hence, it was not possible to determine an exact transition point. The ligated product between the 60 pentapeptide ELP repeat and the sfGFP did not show any transition compared to the pure 60 pentapeptide ELP (data not shown). This concentration dependence is also a well-known characteristic of ELPs.[10]

This PCR-based method can also be employed to change the flanking sequences of the ELP very quickly. Figure 4A shows the underlining principle of the cloning procedure used to install cysteine as an end residue with no cloning scar. Due to the repetitive structure of the ELP gene it was necessary to design primers which anneal at the site of replacement. A *BsaI* recognition loop between ELP annealing and deletion annealing site was necessary to remove the deletion site again afterward. *BsaI* digestion left incompatible 5′ and 3′ sticky ends; therefore, a Klenow fragment was employed to fill the ends. A standard blunt end ligation circularized the linear plasmid (Figure 4A and Supporting Information, Figures S13−S18). This procedure provided an N-terminal cysteine that could be used for bioconjugations to various (macro)molecules (see Supporting Information, DNA sequence 9, protein sequence 5). The cysteine in the ELP is able to form disulfide bonds with different cysteine containing proteins, but also is able to be clicked to other reactive groups like maleimide (i.e., a maleimide-PEG (Figure 4B)). The cloud point determination of Figure 4C shows the influence of PEG conjugation on the ELP cloud point, confirming a shift toward higher temperatures (Figure 4C, CYS-ELP$_{60}$) due to conjugation of the hydrophilic synthetic polymer. However, the same PEG added to a solution of the same ELP that lacked the cysteine functionality did not significantly influence the cloud point (Figure 4C, ELP$_{60}$).

CONCLUSION

The presented approach shows an alternative way to create fast and convenient functional ELPs with sequence lengths up to 600 amino acids, or hundreds of nm in stretched contour length. It allows a straightforward fusion of gene sequences encoding the ELP repeats without any prior vector modifications. We used this approach to demonstrate facile incorporation of functional peptide tags as end groups into ELPs. We demonstrate how this approach was useful for developing end-labeled ELPs through enzyme-mediated site-specific ligation to organic dyes and fluorescent proteins, and show how terminal cysteine incorporation expands the versatile toolbox of bioconjugation opportunities. Since we used a PCR and primer-based approach, our method is essentially sequence independent and does not leave cloning scars. In the future, we anticipate that such a tool for straightforward end-group modification of ELPs will prove useful for developing custom engineered macromolecular systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biomac.5b01726.

Additional information including sequence data (DNA and protein sequences), extended cloning procedures and gel pictures of PCR products and protein purification steps (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.nash@lmu.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from an advanced grant of the European Research Council (Cellufuel Grant 294438), SFB 863, and the Excellence Cluster Center for Integrated Protein Science Munich. M.A.N. acknowledges funding from Society in Science - The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland. We thank the systems biophysics group of Professor Dieter Braun (Ludwig-Maximilians-Universität München) for the access to the JASCO V-650. The authors thank Anna Krautloher for initial concentration determinations of the ELPs at 205 nm. We acknowledge Markus Jobst for his advice on the ongoing manuscript. We thank the following people for providing material to this study: Ellis Durner (eSortase), Angelika Kardinal (TEV protease), and Diana Pippig (Sfp), and Arne Goldenbaum for assistance in lab work. We are grateful for the MALDI-TOF analysis of ELP samples by the protein analysis group of the Ludwig-Maximilians-Universität München (Professor Axel Imhof, Dr. Andreas Schmidt, Dr. Ignasi Forné, and Pierre Schilcher).

REFERENCES

(1) Pack, D. W.; Hoffman, A. S.; Pun, S.; Stayton, P. S. *Nat. Rev. Drug Discovery* **2005**, *4* (7), 581−593.
(2) Onaca, O.; Enea, R.; Hughes, D. W.; Meier, W. *Macromol. Biosci.* **2009**, *9* (2), 129−139.
(3) Nash, M. A.; Waitumbi, J. N.; Hoffman, A. S.; Yager, P.; Stayton, P. S. *ACS Nano* **2012**, *6* (8), 6776−6785.
(4) Nash, M. A.; Gaub, H. E. *ACS Nano* **2012**, *6* (12), 10735−10742.
(5) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. *Philos. Trans. R. Soc., B* **2002**, *357* (1418), 169−184.
(6) Wolff, M.; Braun, D.; Nash, M. A. *Anal. Chem.* **2014**, *86* (14), 6797−6803.
(7) Urry, D. W.; Haynes, B.; Harris, R. D. *Biochem. Biophys. Res. Commun.* **1986**, *141* (2), 749−755.
(8) Tatham, A. S.; Shewry, P. R. *Trends Biochem. Sci.* **2000**, *25* (11), 567−571.
(9) Urry, D. W.; Haynes, B.; Zhang, H.; Harris, R. D.; Prasad, K. U. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (10), 3407−3411.
(10) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2004**, *5* (3), 846−851.
(11) Urry, D. W. *J. Phys. Chem. B* **1997**, *101* (51), 11007−11028.
(12) Bataille, L.; Dieryck, W.; Hocquellet, A.; Cabanne, C.; Bathany, K. *Protein Expression Purif.* **2015**, *110*, 165−171.
(13) Bellucci, J. J.; Amiram, M.; Bhattacharyya, J.; McCafferty, D.; Chilkoti, A. *Angew. Chem., Int. Ed.* **2013**, *52* (13), 3703−3708.
(14) Meyer, D. E.; Chilkoti, A. *Nat. Biotechnol.* **1999**, *17* (11), 1112−1115.
(15) Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater.* **2014**, *10* (4), 1542−1557.
(16) Kojima, C.; Irie, K. *Biopolymers* **2013**, *100* (6), 714−721.
(17) Christensen, T.; Hassouneh, W.; Trabbic-Carlson, K.; Chilkoti, A. *Biomacromolecules* **2013**, *14* (5), 1514−1519.
(18) McDaniel, J. R.; Radford, D. C.; Chilkoti, A. *Biomacromolecules* **2013**, *14* (8), 2866−2872.
(19) Rousseau, R.; Schreiner, E.; Kohlmeyer, A.; Marx, D. *Biophys. J.* **2004**, *86* (3), 1393−1407.
(20) Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D. *Adv. Funct. Mater.* **2015**, *25* (5), 729−738.
(21) Glaves, R.; Baer, M.; Schreiner, E.; Stoll, R.; Marx, D. *ChemPhysChem* **2008**, *9* (18), 2759−2765.

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

(22) Wang, H.; Cai, L.; Paul, A.; Enejder, A.; Heilshorn, S. C. *Biomacromolecules* **2014**, *15* (9), 3421−3428.
(23) Van Eldijk, M. B.; Smits, F. C. M.; Vermue, N.; Debets, M. F.; Schoffelen, S.; Van Hest, J. C. M. *Biomacromolecules* **2014**, *15* (7), 2751−2759.
(24) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3* (2), 357−367.
(25) Amiram, M.; Quiroz, F. G.; Callahan, D. J.; Chilkoti, A. *Nat. Mater.* **2011**, *10* (2), 141−148.
(26) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS One* **2008**, *3* (11), e3647.
(27) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. *Biomacromolecules* **2010**, *11* (4), 944−952.
(28) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. *Nat. Protoc.* **2006**, *1* (1), 280−285.
(29) Huber, M. C.; Schreiber, A.; Wild, W.; Benz, K.; Schiller, S. M. *Biomaterials* **2014**, *35* (31), 8767−8779.
(30) Christensen, T.; Amiram, M.; Dagher, S.; Trabbic-Carlson, K.; Shamji, M. F.; Setton, L. A.; Chilkoti, A. *Protein Sci.* **2009**, *18* (7), 1377−1387.
(31) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One* **2015**, *10* (7), e0131177.
(32) Qi, Y.; Amiram, M.; Gao, W.; McCafferty, D. G.; Chilkoti, A. *Macromol. Rapid Commun.* **2013**, *34*, 1256−1260.
(33) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science* **1999**, *285* (5428), 760−763.
(34) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (44), 15815−15820.
(35) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, C. J.; Hutchison, C. A., III; Smith, H. O. *Nat. Methods* **2009**, *6* (5), 343−347.
(36) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11* (11), 1127−1130.
(37) Robichon, C.; Luo, J.; Causey, T. B.; Benner, J. S.; Samuelson, J. C. *Appl. Environ. Microbiol.* **2011**, *77* (13), 4634−4646.
(38) Studier, F. W. *Protein Expression Purif.* **2005**, *41*, 207−234.
(39) Collins, T.; Azevedo-Silva, J.; da Costa, A.; Branca, F.; Machado, R.; Casal, M. *Microb. Cell Fact.* **2013**, *12* (21), 1−16.
(40) Chow, D. C.; Dreher, M. R.; Trabbic-Carlson, K.; Chilkoti, A. *Biotechnol. Prog.* **2006**, *22* (3), 638−646.
(41) MacEwan, S. R.; Hassouneh, W.; Chilkoti, A. *J. Visualized Exp.* **2014**, *88*, e51583.
(42) Laemmli, U. K. *Nature* **1970**, *227* (5259), 680−685.
(43) Anthis, N. J.; Clore, G. M. *Protein Sci.* **2013**, *22* (6), 851−858.
(44) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (37), 13343−13348.
(45) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. *Proteomics Protocols Handbook* **2005**, 571−607.
(46) Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (5), 1244−1247.
(47) Theile, C.; Witte, M.; Blom, A. *Nat. Protoc.* **2013**, *8* (9), 1800−1807.
(48) Catherine, C.; Oh, S. J.; Lee, K.-H.; Min, S.-E.; Won, J.-I.; Yun, H.; Kim, D.-M. *Biotechnol. Bioprocess Eng.* **2015**, *20* (3), 417−422.
(49) MacKay, J. A.; Callahan, D. J.; FitzGerald, K. N.; Chilkoti, A. *Biomacromolecules* **2010**, *11* (11), 2873−2879.

DOI: 10.1021/acs.biomac.5b01726
Biomacromolecules 2016, 17, 1330−1338

Supporting Information

Sequence Independent Cloning and Post-translational Modification of Repetitive Protein Polymers through Sortase and Sfp-mediated Enzymatic Ligation

Wolfgang Ott[†,‡,§], Thomas Nicolaus[†,‡], Hermann E. Gaub[†,‡], and Michael A. Nash[†,‡,⊥,¶,]*

[†]Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[‡]Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[§]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

[⊥]Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

[¶]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH-Zürich), 4058 Basel, Switzerland.

Original synthesized DNA-Sequence (5' to 3'):

DNA Sequence 1:

GTACCAGGCGTTGGTGTGCCGGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGTG
GGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGGC
GGCGGTGTTCCGGTTGGTGGCGTTCCGGGTGGCGGT

Translated Protein-Sequence (N-Terminus to C-Terminus):

Protein Sequence 1:

VPGVGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPVGGVPGGG

PCR modified DNA-Sequence (5' to 3'):

DNA Sequence 2:

GTGCCGGGAGAAGGAGTCCCTGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGT
GGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGG
CGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGA

Translated Protein-Sequence (N-Terminus to C-Terminus):

Protein Sequence 2:

VPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEG

Sequence Maps were built with SnapGene 3.0.3 (GSL Biotech LLC, Chicago, IL, USA)

Sequence of original pET28a-Vector with aligned Primers:

Forward and reverse primer flanking the multiple cloning site of the pET28a vector, and deleting the restriction sites, HIS-tags as well as the thrombin site and the T7-tag (Fig. S1).

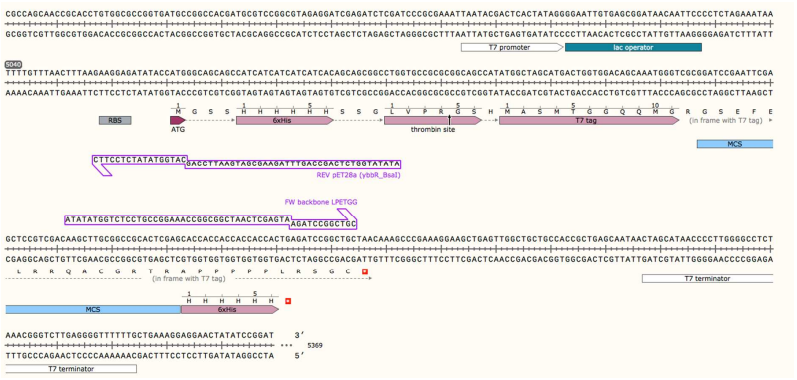


Figure S1. MCS of pET28a vector with two primers suitable for linearization and insertion of the ybbR- and sortase c-tag. The linear vector serves as starting template for the ELP insertion.

DNA Sequence (5' to 3'):

Colored letters represent the annealing region of the forward (green) and reverse (red) primer.

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAA
TTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCAT
CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
GCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTC
GAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGC
TGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAA
ACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

FW backbone LPETGG (Primer 5, Table 1, Main text)

ATATATGGTCTCCTGCCGGAAACCGGCGGCTAACTCGAGTAAGATCCGGCTGC

REV pET28a (ybbR_BsaI) (Primer 12) - *Theoretical primer, not used in this study*:

ATATATGGTCTCAGCCAGTTTAGAAGCGATGAATTCCAGCATGGTATATCTCCTTC

Sequence of the modified pET28a-vector already containing a ybbR-tag:

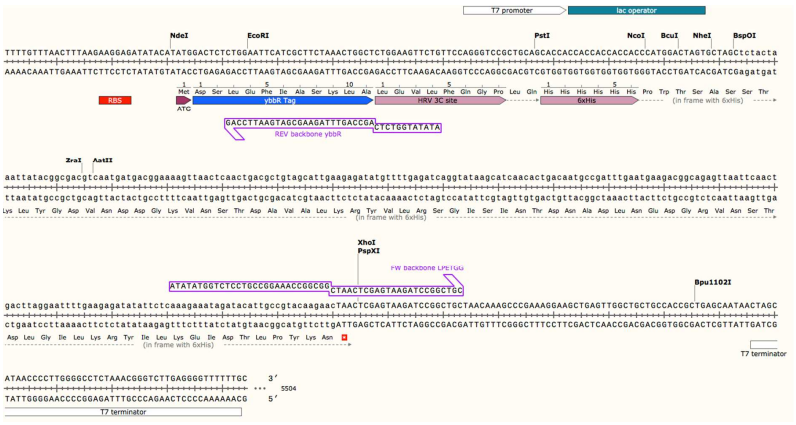


Figure S2. MCS of modified pET28a vector used in this study already having a ybbR-tag. The linear vector serves as starting template for the ELP insertion.

DNA Sequence (5' to 3'):

Colored letters represent the annealing region of the forward (green) and reverse (red) primers.

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTGTAGAAATA
ATTTTGTTTAACTTTAAGAAGGAGATATACATATGGACTCTCTGGAATTCATCGCTTC
TAAACTGGCTCTGGAAGTTCTGTTCCAGGGTCCGCTGCAGCACCACCACCACCACCA
CCCATGGACTAGTGCTAGCTCTACTAAATTATACGGCGACGTCAATGATGACGGAAA
AGTTAACTCAACTGACGCTGTAGCATTGAAGAGATATGTTTTGAGATCAGGTATAAG
CATCAACACTGACAATGCCGATTTGAATGAAGACGGCAGAGTTAATTCAACTGACTT
AGGAATTTTGAAGAGATATATTCTCAAAGAAATAGATACATTGCCGTACAAGAACT
AACTCGAGTAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT
GCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAG
GGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

FW backbone LPETGG (Primer 5, Table 1, Main text)

ATATATGGTCTCCTGCCGGAAACCGGCGGCTAACTCGAGTAAGATCCGGCTGC

REV backbone ybbR (Primer 6, Table 1, Main text)

ATATATGGTCTCAGCCAGTTTAGAAGCGATGAATTCCAG

Cloning Site of Vector A (Fig. 1):

Figure S3. Top: Illustration of the linearized plasmid after PCR amplification, Middle: Zoom in of the *Bsa*I-digested 5'-end, Bottom: Zoom in of the *Bsa*I-digested 3'-end.

PCR Product ELP I (5' to 3'):

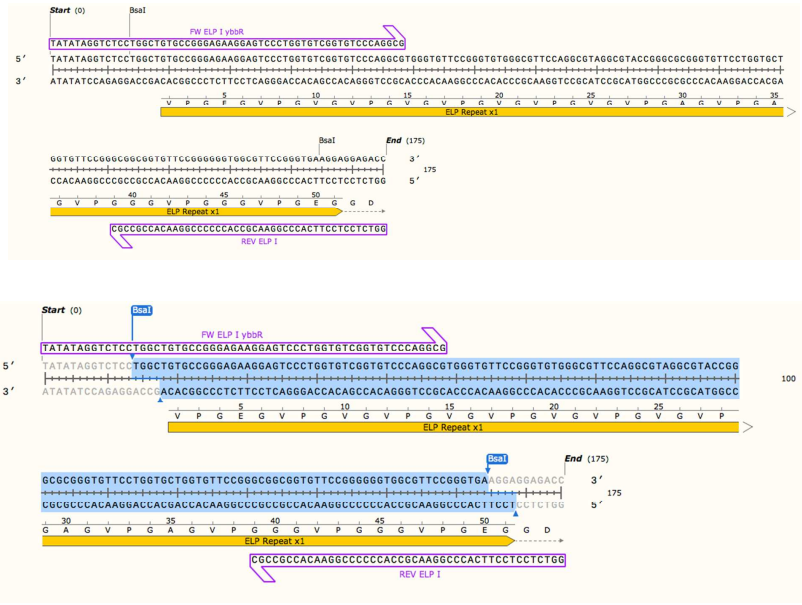


Figure S4. PCR Product "ELP I" (top) with the two corresponding primers (FW ELP I (Primer 1a, Table 1, Main text) ybbR, REV ELP I (Primer 1b, Table 1, Main text)). The bottom figure shows the *Bsa*I-digested fragment.

DNA Sequence of ELP I:

DNA Sequence 3:

TATATAGGTCTCCTGGCTGTGCCGGGAGAAGGAGTCCCTGGTGTCGGTGTCCCAGGC
GTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCT
GGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGAGG
AGACC

PCR Product ELP II (5' to 3'):

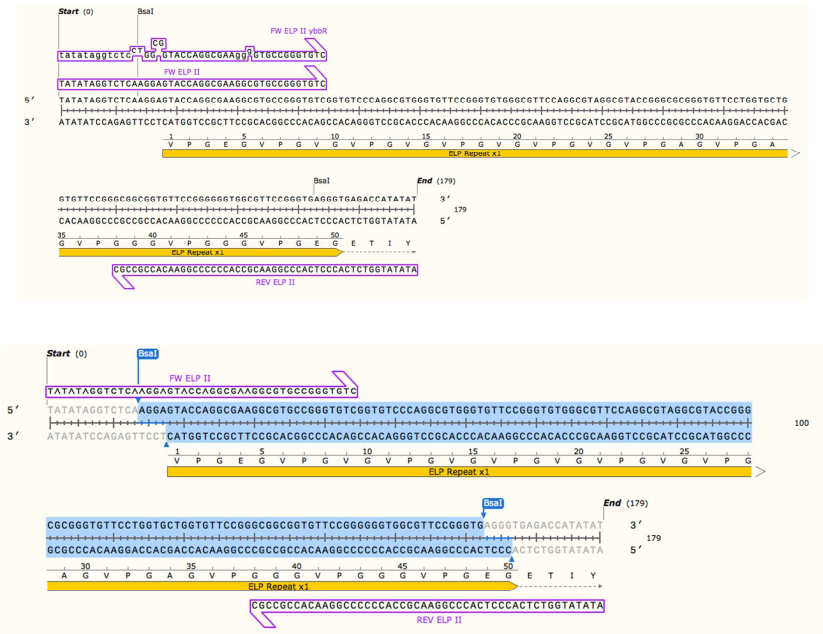


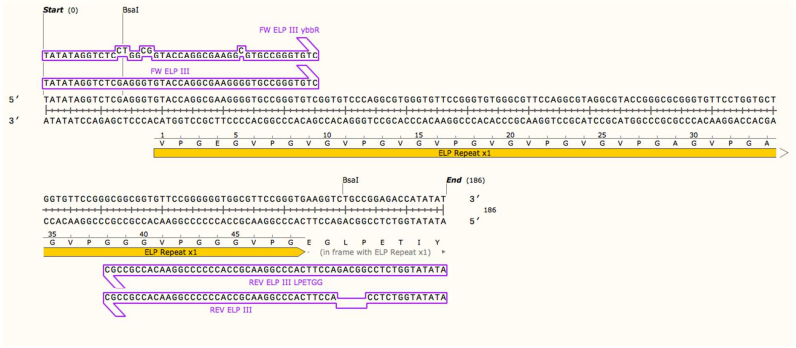
Figure S5. PCR Product "ELP II" (top) with the two corresponding primers (FW ELP II (Primer 2a, Table 1, Main text), REV ELP II (Primer 2b, Table 1, Main text)). FW ELP II ybbR (Primer 8, Table 1, Main text) would be necessary if the ELP II fragment should be ligated first to the ybbR-tag on the linearized backbone instead of the fragment ELP I. The bottom figure shows the *Bsa*I-digested fragment.

DNA Sequence of ELP II:

DNA Sequence 4:

TATATAGGTCTCAAGGAGTACCAGGCGAAGGCGTGCCGGGTGTCGGTGTCCCAGGC
GTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCT
GGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGGTGGCGTTCCGGGTGAGGGTGA
GACCATATAT

PCR Product ELP III (5' to 3'):



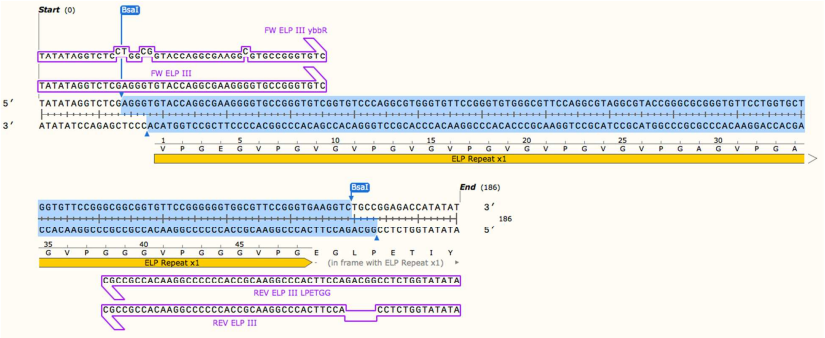


Figure S6. PCR Product "ELP III" (top) with the two corresponding primers (FW ELP III (Primer 3a, Table 1, Main text), REV ELP LPETGG (Primer 3b, Table 1, Main text)). FW ELP III ybbR (Primer 9, Table 1, Main text) would be necessary if the ELP III fragment is to be ligated first to the ybbR-tag on the linearized backbone instead of the fragment ELP I. REV ELP III (Primer 4, Table 1, Main text) is necessary to create a compatible sticky end for the ELP growing reaction after inserting the first three ELP fragments (I, II, III). The bottom figure shows the *Bsa*I-digested fragment.

DNA Sequence of ELP III:

DNA Sequence 5:

TATATAGGTCTCGAGGGTGTACCAGGCGAAGGGGTGCCGGGTGTCGGTGTCCCAGG
CGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCC
TGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGTCT
GCCGGAGACCATATAT

PCR Product ELP IV (5' to 3'):

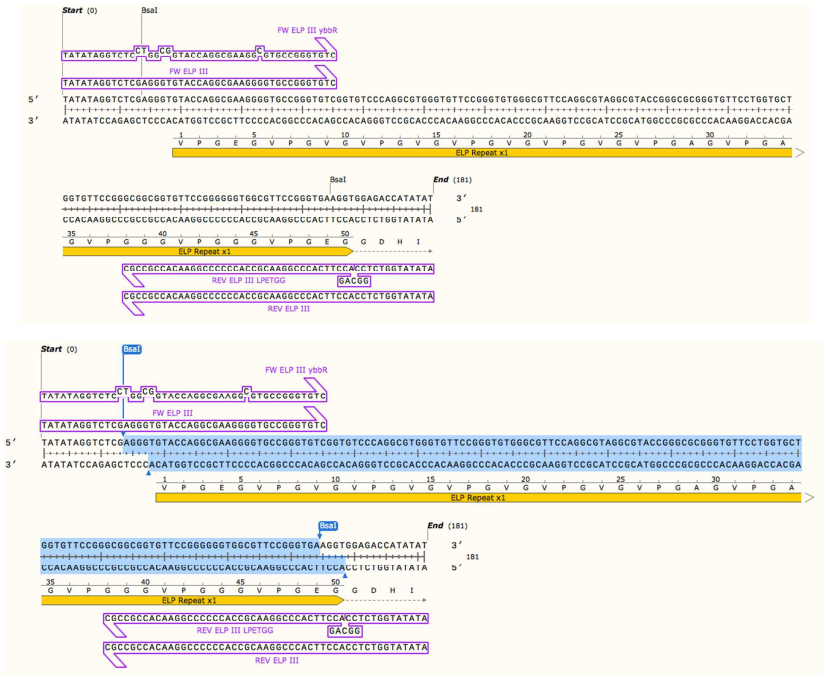


Figure S7. PCR Product "ELP IV" (top) with the two corresponding primers (FW ELP III (Primer 3a, Table 1, Main text), REV ELP III (Primer 4, Table 1, Main text)). FW ELP III ybbR (Primer 9, Table 1, Main text) would be necessary if the ELP III fragment is to be ligated first to the ybbR-tag on the linearized backbone instead of the fragment ELP I. REV ELP III LPETGG (Primer 3b, Table 1, Main text) is necessary to create a compatible sticky end for the initial ELP insertion with fragments ELP I, II and III. ELP IV is necessary for the expansion reaction of the ELP insert (ELP IV replaces ELP III). The bottom figure shows the *Bsa*I-digested fragment.

DNA Sequence of ELP IV:

DNA Sequence 6:

TATATAGGTCTCGAGGGTGTACCAGGCGAAGGGGTGCCGGGTGTCGGTGTCCCAGG
CGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCC
TGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGTG
GAGACCATATAT

Insertion site for ELP plasmid (3x 10 pentapepitdes upwards)

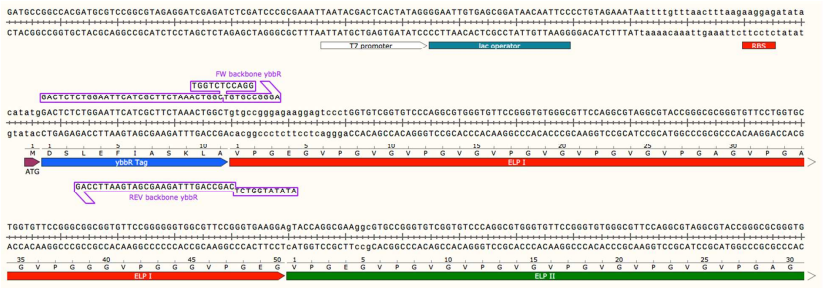


Figure S8. Illustrating the insertion site for the plasmid linearization reaction. FW backbone ybbR (Primer 7, Table 1, Main text) and REV backbone ybbR (Primer 6, Table 1, Main text) open the plasmid right after the N-terminal ybbR-tag and the first following ELP fragment I. The three following ELP fragments I, II and IV are ligated in between.

Linearized PCR product:

The construct now has a ybbR-tag at the 5'- and 3'- end and gets cleaved off after *Bsa*I digestion at the 5' end, leaving an ELP I sticky 5' end for the Golden Gate reaction and a ybbR-sticky end at the 3' end. For the linearization reactions of growing ELP constructs the insertion site always remains the same. This means that the structure of the linear backbone is:

- 5' ybbR-[ELP I-ELP II-ELP III]$_n$-LPETGG-Backbone-ybbR 3'
- *Bsa*I Digestion: 5'-[ELP I-ELP II-ELP III]$_n$-LPETGG-Backbone-ybbR 3'
- Inserts are added between the sticky ends of the ybbR-tag and the first ELP I fragment.

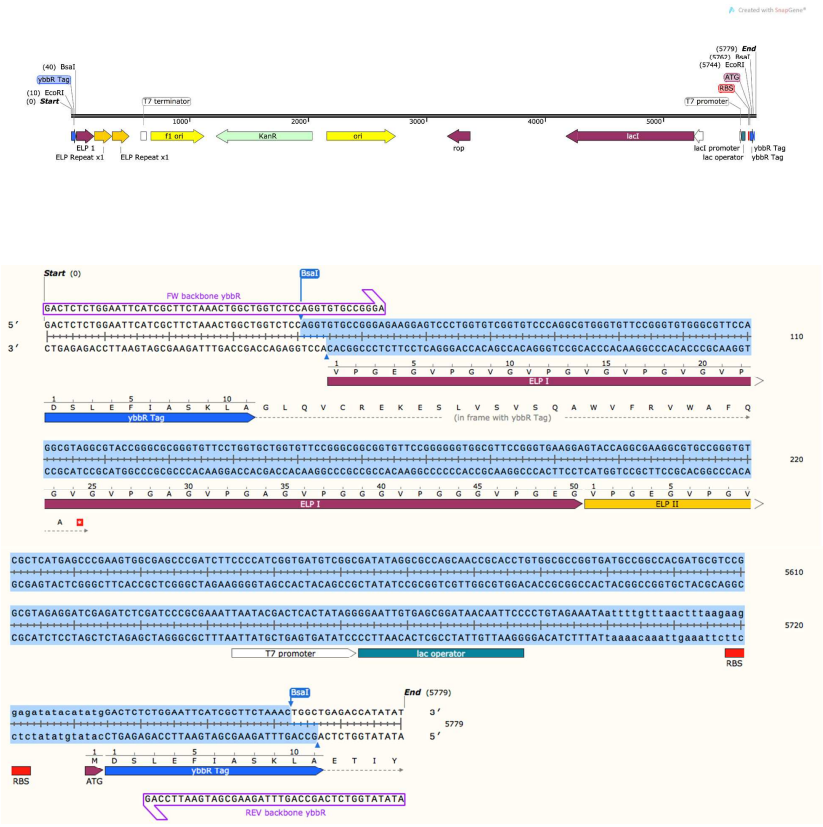


Figure S9. Top: Illustration of the linearized plasmid after PCR amplification with the primers "FW backbone ybbR (Primer 7, Table 1, Main text) and REV backbone ybbR (Primer 6, Table 1, Main text) and REV backbone ybbR", Middle: Zoom in of the *Bsa*I-digested 5'-end, Bottom: Zoom in of the *Bsa*I-digested 3'-end.

Cloning efficiencies:

Table S1. Overview of cloning efficiencies.

Pentapeptide Repeats	Number of colonies analyzed with colony PCR/ restriction digestion (*)	Number of clones with correct size in colony PCR/ restriction digestion	Number of sequencing reactions performed	Correct clones (#)	Ratio
10	14	6	2	2	100.00%
20	14	6	3	2	66.67%
30	14	4	1	1	100.00%
40	14	1	1	1	100.00%
50	14	3	3	2	66.67%
60	14	3	3	2	66.67%
70	14	4	1	1	100.00%
80	14	4	1	1	100.00%
90	14	4	1	1	100.00%
120	14	2	2	1	50.00%
60 with Cysteine	6	6	2	2	100.00 %

(*) Constructs longer than 60 pentapeptide repeats were analyzed by restriction digestion.
(#) Wrong results missed 1x ELP Fragment (10 Pentapeptides)

Gel pictures of the PCR amplified ELP fragments I-IV and the different linearized plasmids used in this study:

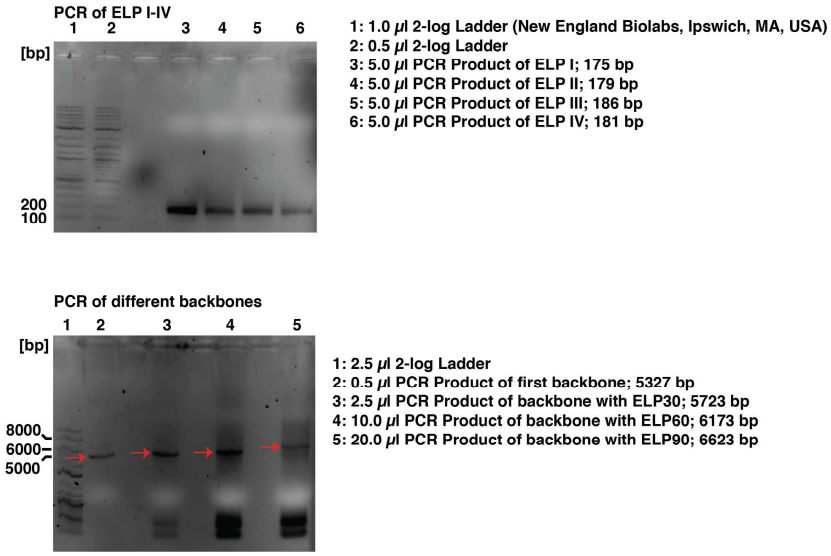


Figure S10. Gel pictures showing PCR products of insert (top) and linearized plasmid (bottom). Bands with a red arrow are the corresponding linearized plasmids and were isolated via gel extraction to separate them from the side products.

Exemplary sequence of ybbR-ELP$_{30}$-LPETGG:

ybbR-Tag
ELP 1
ELP 2
ELP 3
Sortase c-tag (LPETGG)

DNA Sequence 7 (5' to 3'):

ATGGACTCTCTGGAATTCATCGCTTCTAAACTGGCTGTGCCGGGAGAAGGAGTCCCT
GGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTA
CCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGC
GTTCCGGGTGAAGGAGTACCAGGCGAAGGCGTGCCGGGTGTCGGTGTCCCAGGCGT
GGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGG
TGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAGGGTGTACC
AGGCGAAGGGGTGCCGGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGTGGGCG
TTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGGCGGCG
GTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGTCTGCCGGAAACCGGCGGCTAA

Protein Sequence 3 (N- to C-terminus):

MDSLEFIASKLAVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGG
VPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVP
GEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGLPETGG

Growing ELP constructs had repeats of the ELP 1, 2 and 3 segments inserted in between the ybbR-tag and the following ELP sequence.

Protein purification and Sfp-mediated labeling: gels with samples from different steps:

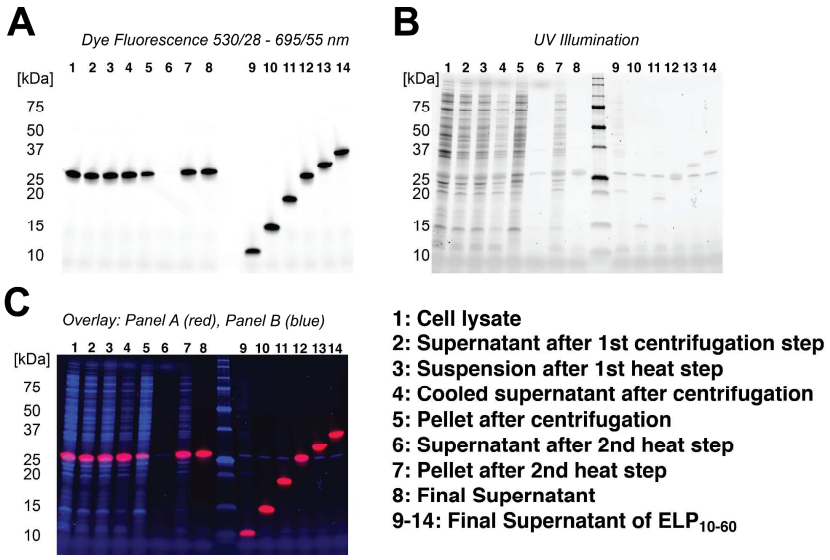
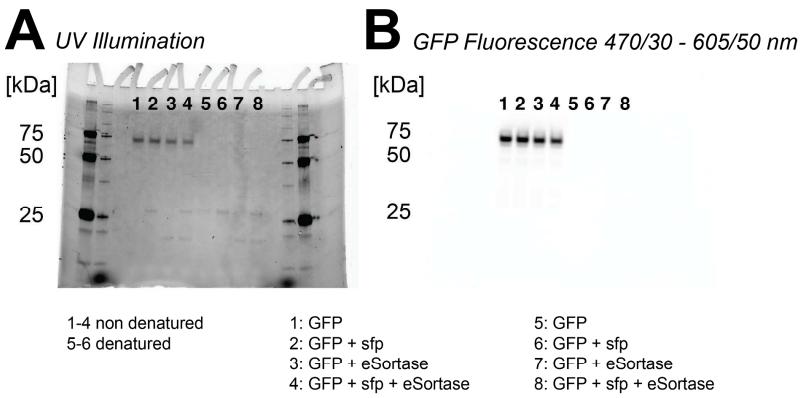


Figure S11. Gel of purification steps of the ELP_{40} (lane 1-8) via ITC with different illumination methods. Lane 9-14 shows the purified ELP_{10-60} constructs. Sfp appears at about 25 kDa throughout all lanes. Subset A shows the fluorescence of the Sfp catalyzed labeling reaction of CoA-647 dye to the ybbR-tag containing ELP constructs. Panel B shows the UV activated tryptophan labeling (ELPs not labeled) via Bio-Rad Stain-Free™ technology. C shows the overlay of A (red) and B (blue).

Negative control confirming native GG-sfGFP on top of SDS-Gel (Figure 2)



1-4 non denatured
5-6 denatured

1: GFP
2: GFP + sfp
3: GFP + eSortase
4: GFP + sfp + eSortase

5: GFP
6: GFP + sfp
7: GFP + eSortase
8: GFP + sfp + eSortase

Figure S12. Negative control for Figure 2 B-D, showing the electrophoresis running behavior of native GG-sfGFP. Sortase in lane 3, 4, 7, and 8 at ca. 18 kDa. Sfp in lane 2, 4, 6 and 8 at ca. 25 kDa. GG-sfGFP (native) in lane 1-4 at ca 75 kDa and GG-sfGFP (denatured) lane 5-8 at ca. 25 kDa.

Sequence of HIS-TEV-GG-sfGFP:

HIS_6-Tag
TEV-Site
Glycines
sfGFP

DNA Sequence 8 (5' to 3'):

ATGCACCACCACCACCACCACGGAGAAAACCTGTACTTCCAGGGAGGGGGCGGTAG
CAAAGGTGAAGAACTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGTGA
TGTTAATGGCCACAAATTTTCAGTTCGTGGTGAAGGCGAAGGTGATGCAACCATTGG
TAAACTGACCCTGAAATTTATCTGTACCACCGGCAAACTGCCGGTTCCGTGGCCGAC
CCTGGTTACCACCCTGACCTATGGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATG
AAACGCCACGATTTTTTCAAAAGCGCAATGCCGGAAGGTTATGTTCAAGAACGTACC
ATCTCCTTTAAAGACGACGGTAAATACAAAACCCGTGCCGTTGTTAAATTTGAAGGT
GATACCCTGGTGAATCGCATTGAACTGAAAGGCACCGATTTTAAAGAGGATGGTAA
TATCCTGGGCCACAAACTGGAATATAATTTCAATAGCCACAACGTGTATATCACCGC
AGACAAACAGAAAAATGGCATCAAAGCCAATTTTACCGTGCGCCATAATGTTGAAG
ATGGTAGCGTGCAGCTGGCAGATCATTATCAGCAGAATACCCCGATTGGTGATGGTC
CGGTTCTGCTGCCGGATAATCATTATCTGAGCACCCAGACCGTTCTGAGCAAAGATC
CGAATGAAAAACGTGATCATATGGTGCTGCATGAGTATGTTAATGCAGCAGGTATTA
CCCATGGTATGGATGAGCTGTATAAGTAA

Protein Sequence 4 (N- to C-terminus):

MHHHHHHGENLYFQGGGGSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGK
LTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISF
KDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQK
NGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRD
HMVLHEYVNAAGITHGMDELYK

After TEV digestion:

GGGGSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKF
EGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITH
GMDELYK

Gel picture of the PCR amplified, linearized and mutated Cysteine-ELP_{60}-LPETGG:

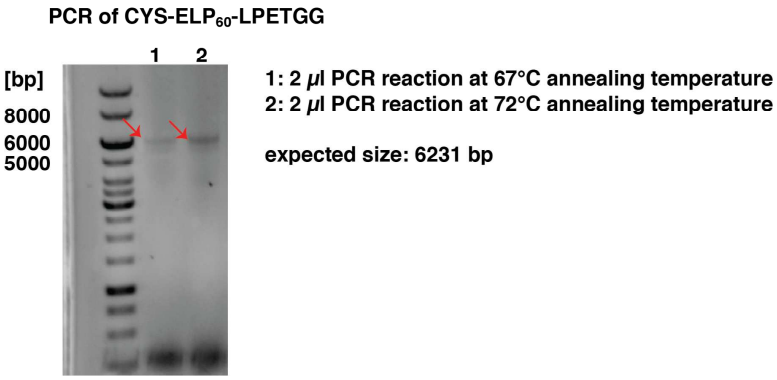


Figure S13. Gel picture showing the PCR products of the linearized and mutated Cysteine-ELP_{60}-LPETGG bands at 67 and 72°C annealing temperature. The red arrows indicate the desired PCR product.

PCR product (5' to 3'):

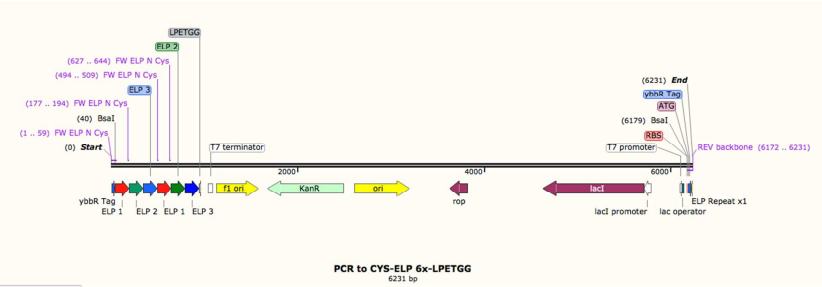


Figure S14. Illustration of PCR product before *Bsa*I restriction. The plasmid was linearized with the primers "FW ELP N Cys (Primer 10, Table 1, Main text) and REV backbone (Primer 11, Table 1, Main text)".

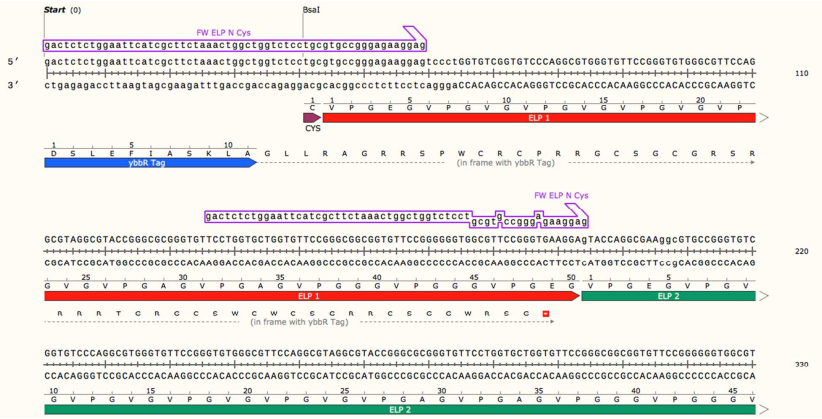


Figure S15. Zoom in 5' end of PCR product before *Bsa*I restriction.

Figure S16. Zoom in 3' end of PCR product before *Bsa*I restriction.

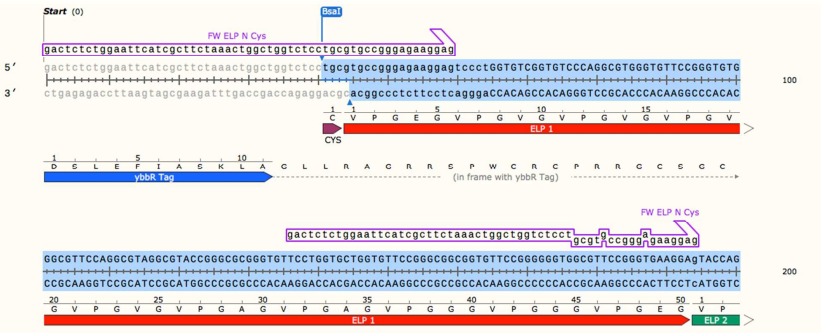


Figure S17. Zoom in 5' end of PCR product after *Bsa*I restriction.

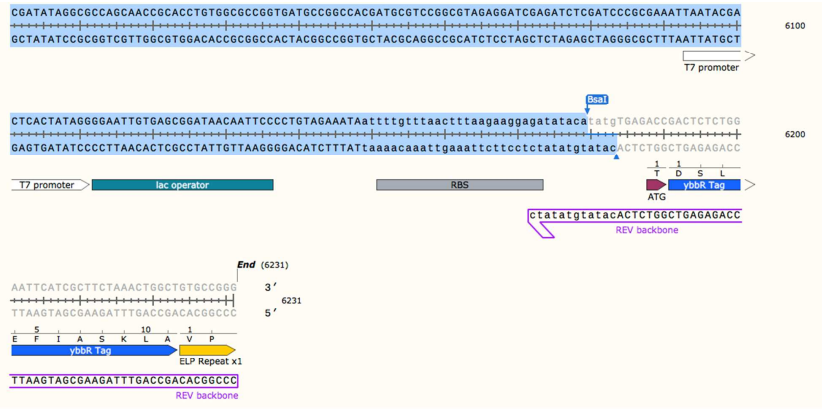


Figure S18. Zoom in 3' end of PCR product after *Bsa*I restriction.

The digested ends were filled to blunt ends with Klenow Fragment and ligated with the T4-Ligase as described in the main text (methods).

Sequence of final Cysteine-ELP$_{60}$-LPETGG:

Cysteine
ELP 1
ELP 2
ELP 3
Sortase c-tag (LPETGG)

DNA Sequence 9 (5' to 3'):

ATGTGCGTGCCGGGAGAAGGAGTCCCTGGTGTCGGTGTCCCAGGCGTGGGTGTTCCG
GGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTT
CCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGAGTACCAGGCGAAGG
CGTGCCGGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGT
AGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGG
GGGTGGCGTTCCGGGTGAGGGTGTACCAGGCGAAGGGGTGCCGGGTGTCGGTGTCC
CAGGCGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGT
GTTCCTGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAA
GGTGTGCCGGGAGAAGGAGTCCCTGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGG
TGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCC
GGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGGAGTACCAGGCGAAGGCG
TGCCGGGTGTCGGTGTCCCAGGCGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAG
GCGTACCGGGCGCGGGTGTTCCTGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGG
GTGGCGTTCCGGGTGAGGGTGTACCAGGCGAAGGGGTGCCGGGTGTCGGTGTCCCA
GGCGTGGGTGTTCCGGGTGTGGGCGTTCCAGGCGTAGGCGTACCGGGCGCGGGTGT
TCCTGGTGCTGGTGTTCCGGGCGGCGGTGTTCCGGGGGGTGGCGTTCCGGGTGAAGG
TCTGCCGGAAACCGGCGGCTAA

Protein Sequence 5 (N- to C-terminus):

MCVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGE
GVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVG
VPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVP
GVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPG
VGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGA
GVPGAGVPGGGVPGGGVPGEGLPETGG

MALDI-TOF analysis of ELP_{30}, ELP_{40} and ELP_{50}

For further analysis, three ELP samples were sent to MALDI-TOF analysis to check their grade of purity.

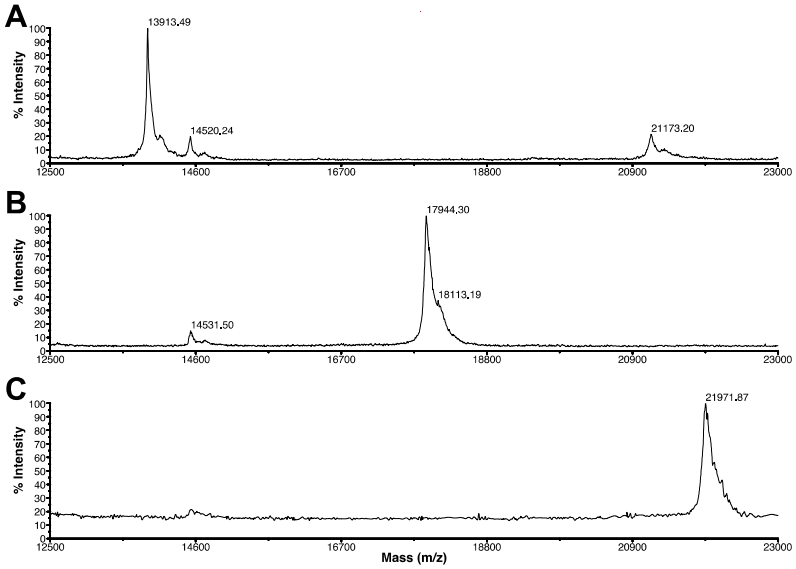


Figure S19. MALDI-TOF Analysis of ELP_{30} (A, 13914 Da), ELP_{40} (B, 17926 Da) and ELP_{50} (C, 21974 Da).

SDS-Gel of Bioconjugation Reactions:

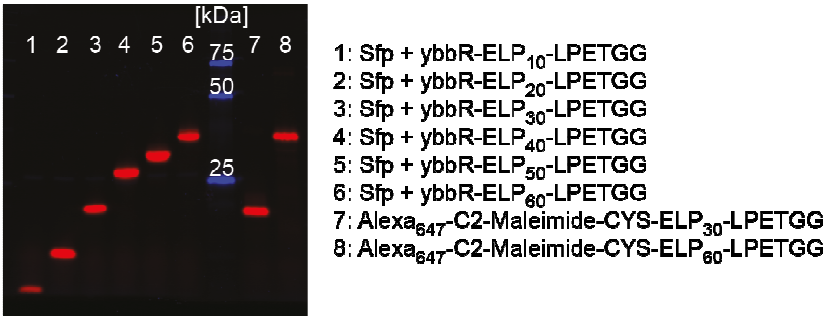


Figure S20. An SDS gel picture confirming Sfp-mediated ELP conjugation to a CoA-647 dye (Lanes 1-6; ELP_{10}-ELP_{60}). Lanes 7 and 8 of the gel show the result of the $Alexa_{647}$-maleimide-dye bioconjugation to the cysteine-ELP_{30} (7) and cysteine-ELP_{60} (8). It was not possible to label the cysteine-ELPs with the CoA-647 dye, since it did not carry a ybbR-tag anymore. Vice versa, the $Alexa_{647}$-maleimide dye did not label the standard ybbR-containing ELPs, which do not have a cysteine (data not shown). Red is the dye fluorescence (excitation: 530/28 nm and emission: 695/55 nm). The marker lane is shown in blue.

6.3 Associated Publication P8

Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy

by

Wolfgang Ott*, Markus A. Jobst*, Magnus S. Bauer, Ellis Durner,

Lukas F. Milles, Michael A. Nash, and Hermann E. Gaub

*(*contributed equally)*

published in

ACS Nano, 11 (6), 6346-6354, (2017)

Reprinted from [100], with permission from American Chemical Society.

ACSNANO www.acsnano.org

ARTICLE

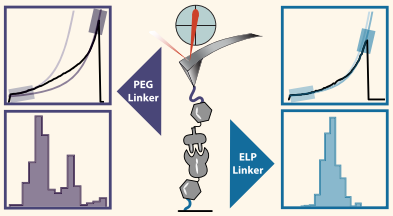
Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy

Wolfgang Ott,[†,‡,⊥] Markus A. Jobst,[†,⊥] Magnus S. Bauer,[†] Ellis Durner,[†] Lukas F. Milles,[†] Michael A. Nash,[§,∥] and Hermann E. Gaub[*,†]

[†]Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany
[‡]Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[§]Department of Chemistry, University of Basel, 4056 Basel, Switzerland
[∥]Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland

S *Supporting Information*

ABSTRACT: **Single-molecule force spectroscopy (SMFS) is by now well established as a standard technique in biophysics and mechanobiology. In recent years, the technique has benefitted greatly from new approaches to bioconjugation of proteins to surfaces. Indeed, optimized immobilization strategies for biomolecules and refined purification schemes are being steadily adapted and improved, which in turn has enhanced data quality. In many previously reported SMFS studies, poly(ethylene glycol) (PEG) was used to anchor molecules of interest to surfaces and/or cantilever tips. The limitation, however, is that PEG exhibits a well-known trans−trans−gauche to all-trans transition, which results in marked deviation from standard polymer elasticity models such as the worm-like chain, particularly at elevated forces. As a result, the assignment of unfolding events to protein domains based on their corresponding amino acid chain lengths is significantly obscured. Here, we provide a solution to this problem by implementing unstructured elastin-like polypeptides as linkers to replace PEG. We investigate the suitability of tailored elastin-like polypeptides linkers and perform direct comparisons to PEG, focusing on attributes that are critical for single-molecule force experiments such as linker length, monodispersity, and bioorthogonal conjugation tags. Our results demonstrate that by avoiding the ambiguous elastic response of mixed PEG/peptide systems and instead building the molecular mechanical systems with only a single bond type with uniform elastic properties, we improve data quality and facilitate data analysis and interpretation in force spectroscopy experiments. The use of all-peptide linkers allows alternative approaches for precisely defining elastic properties of proteins linked to surfaces.**



KEYWORDS: *single-molecule force spectroscopy, elastin-like polypeptides, biopolymer spacer, sortase coupling, protein ligation*

Refined Techniques in SMFS. Single-molecule force spectroscopy (SMFS) is a state-of-the-art technique in the rapidly growing field of molecular biomechanics.[1−3] Tools and methods are being steadily developed to improve ease of sample handling, sensitivity, reproducibility, and reliability.[4,5] In parallel, the biochemical toolbox is expanded continuously, enabling analysis of more complex and demanding biological systems. Improvements such as the use of orthogonal binding handles,[6−9] diverse biomolecule immobilization strategies,[10−14] and alternative methods for protein synthesis (*i.e.*, recombinant bulk expression or cell-free *in vitro* expression) are all examples of significant technical advances that have been achieved in recent years.[15]

Requirements for Recording Large Data Sets and Challenges Arising Therefrom. A key requirement to probe multiple different protein domains in a single experiment is the ability to use a single cantilever over extended periods of time to achieve a large number of force−extension traces. For this purpose, two main advances are worth noting, the first of them being the improvement of geometrically defined covalent surface tethering and the second being the discovery and characterization of the type III cohesin−dockerin (Coh:Doc) interaction.[7] Coh:Doc receptor−ligand pairs can withstand remarkably high forces in a SMFS assays and exhibit extremely high long-term functionality. This latter property is particularly important for carrying out multiplexed experiments where many proteins deposited onto the same surface and spatially

Received: April 18, 2017
Accepted: June 7, 2017
Published: June 7, 2017

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

separated are pulled apart using the same receptor-modified cantilever. In such a configuration, Coh:Doc is used as a binding handle to successfully and continuously unfold target proteins for over 24 h of measurement time without significant loss of binding activity. Data sets of typically several tens of thousands of force−extension curves can easily be obtained using type III Coh:Doc, dramatically outperforming other mechanostable interactions (*e.g.*, biotin−avidin).

The ability to measure with a single cantilever over several days allows interrogation of different types or variants of proteins immobilized on different positions of the same substrate (*i.e.*, protein microarrays) and to achieve statistical significance over the course of a single experiment. This leads to large data sets and requires the use of sophisticated algorithms to identify and extract specific single-molecule interactions among a large number of traces with poor signal, such as empty traces, multiple interactions in parallel, or nonspecific interactions. Independent of the size of the data sets though, elasticity models whether applied as part of elaborate algorithms or fitted manually to single curves have in the past been required to account for the different elastic contributions stemming from heterogeneous stretching behavior of mixed poly(ethylene glycol) (PEG)−protein polymer backbone.

Conformational Changes of PEG Linker Molecules Obscure Molecular Characteristics of Interest. When performing SMFS in an elevated force regime using PEG as linker molecules, additional challenges arise. A conformational transition of PEG occurs in a force range of up to *ca.* 300 pN, resulting in an approximately linear force−extension regime.[16−18] In aqueous solutions, PEG exhibits a trans−trans−gauche conformation. With rising force on the polymer, the occupancy of conformations is shifted to all-trans, effectively increasing the net polymer contour length. Analysis methods such as fitting standard elasticity models to the data or detecting contour length increments within said force range are therefore compromised and would, for a quantitative description, require improved heterogeneous elasticity models.

PEG is a highly flexible polymer with a low persistence length, while peptide bonds have restricted degrees of freedom. These restrictions alter the stretching behavior and give rise to marked differences in comparison to PEG. Furthermore, the ratio of PEG linker length to unfolded protein backbone length is not constant over the course of an unfolding trace, which means fitting parameters must be optimized for different sections of the curve as more domains unfold. This issue becomes particularly significant and noticeable when probing protein unfolding and receptor−ligand unbinding in a high force regime and is also problematic when unfolding occurs across a broad range of forces.

Benefits of ELP Linkers in SMFS. In this study we investigate the feasibility of biological peptide polymers to circumvent this problem. We selected well-characterized elastin-like polypeptides (ELPs) as a suitable candidate for this purpose. The progression of cloning techniques of repetitive genes in recent years has set the stage for precisely defined protein polymers and opened up the ability to design, produce, and purify protein spacers of well-defined contour length and chemical composition for single-molecule experiments.[19−22] ELPs exhibit similar elasticity behavior as unfolded protein backbone and are completely monodisperse, a key advantage compared to synthetic polymers such as PEG. Monodisperse ELP linkers fused directly to a protein of interest allow for complete control of the lengths of a nanomechanical system from the surface up to the force transducer, which is not true for the chemically synthesized PEG polymers with non-negligible polydispersity. Since ELPs are expressed recombinantly in *Escherichia coli* (*E. coli*), their production is easily scaled up, resulting in lower costs compared to commercially available heterobifunctional PEGs. Furthermore, ELPs can be produced with N-/C-terminal protein ligation tags, which can be used for specific and bio-orthogonal surface chemistry in SMFS sample preparation.

ELPs are synthetic biopolymers derived from tropoelastin domains. They are composed of a repetitive amino acid heptamer "Val-Pro-Gly-Xaa-Gly",[23] where Xaa is a guest residue that can be any amino acid apart from proline. The guest residue influences the hydrophobicity of the protein and impacts the lower critical solution temperature, the point at which the ELP undergoes a soluble-to-insoluble phase transition. At this environment-dependent cloud point, ELPs change their conformation and precipitate, resulting in clouding of the solution.

ELPs are intrinsically disordered proteins that do not fold into well-defined secondary and tertiary structures, but rather remain unfolded and flexible, a property that is ideally suited to their application as spacer/linker molecules for SMFS.[24] We hypothesized that ELPs would therefore be a suitable choice to achieve both surface passivation and site-specific immobilization in single-molecule nanomechanical experiments. The bulky yet flexible features of ELPs inhibit nonspecific protein binding to the surface, while enabling ligation of other proteins due to the high degree of accessibility of N- or C-terminally fused peptide tags. Post-translational protein ligation methods have made it possible to move from organic chemical conjugation methods toward enzyme-mediated covalent immobilization, for example utilizing sortase A or Sfp.[14,25] Both enzymes catalyze sequence- and site-specific reactions yielding uniform protein orientation at the surface.

ELPs have previously been the subject of atomic force microscopy (AFM) studies. For example, AFM was used to support theoretical predictions about the behavior of ELPs above and below their cloud point, as well as to study ELP elasticity.[26−28] This study was carried out entirely below the cloud point, so that intermolecular interactions between ELPs were negligible. In contrast to prior studies, we employ ELPs as spacer molecules with other protein domains attached. Our results show that ELPs provide several benefits over PEG linkers in SMFS attributable primarily to the features of having uniform elastic properties and monodisperse linkers.

This study offers an attractive substitute for established PEG systems using all-protein ELP linkers. The immobilization strategy provides precise control over the elastic properties of multicomponent protein mechanical systems linked between a glass surface and a force transducer. Our approach transfers advances in smart polymer research to SMFS experiments and describes the improvements achieved through this alternative surface anchoring strategy.

RESULTS AND DISCUSSION

SMFS with Receptor−Ligand Polyproteins Employing Site-Specific Immobilization. Typically PEG linkers with an *N*-hydroxysuccinimide (NHS) group are linked to an amino-silanized surface. The other end of the PEG contains a reactive group for protein immobilization, which in most cases is a thiol-reactive maleimide group. Figure 1A illustrates a Coh:Doc-

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

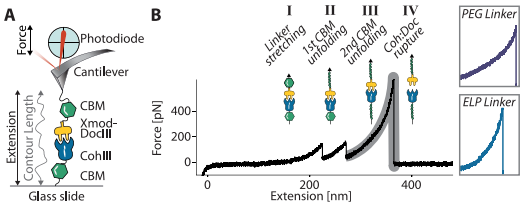


Figure 1. (A) SMFS configuration: Cantilevers are functionalized with CBM-Xmod-DocIII fusion proteins. Glass slides are modified with CohIII-CBM constructs. (B) Coh:Doc-based SMFS unfolding trace. Following Coh:Doc complex formation at zero extension, retraction of the cantilever results in mechanical stretching of the receptor:ligand-linked polyprotein. (I) Spacer molecules are fully extended and stretched. (II, III) The weakest links in the chain, usually the fingerprint domains (here: CBM), are unfolded in series. (IV) Finally, the Coh:Doc complex dissociates under force. The unfolded CBM domains can then refold after the complex rupture. The cantilever is now free to probe a different molecule on the surface. The insets on the right side qualitatively illustrate the differences in linker stretching in the high-force regime as observed in the final peak for constructs immobilized using PEG and ELP linkers. A quasi-linear regime of PEG stretching attributable to the conformational transition from trans-trans-gauche to all-trans is clearly visible for PEG in contrast to ELP.

based SMFS experiment. Proteins anchored to a functionalized glass surface are probed by the corresponding receptor fusion protein covalently linked to the cantilever tip. A characteristic unfolding curve recorded at constant speed is shown in Figure 1B. After the Coh:Doc complex is formed by contacting the cantilever with the surface, force is applied by retracting the base of the cantilever. The signal is detected by a quadrant photodiode with a laser that is reflected off the back side of the cantilever. Bending of the cantilever is translated into a differential voltage output of the photodiode. Upon retraction of the cantilever base at constant speed, the polymer linker is stretched first (Figure 1B, I). Subsequently, the weakest component in the system unfolds. In this case two carbohydrate binding modules (CBMs) are unfolded consecutively (Figure 1B, II and III). Finally, the force increases to a level where the receptor ligand pair dissociates. Following Coh:Doc rupture, the force drops to zero (Figure 1B, IV) and the cantilever is free to probe another molecule at a different location on the surface.

In order to identify data traces that show specific single-molecule interactions, a multilevel sorting algorithm is used to search for characteristic unfolding patterns of the fingerprint domains. This algorithm takes into account the unfolding forces and the measured increases in contour length (*i.e.*, contour length increments) of the peptide backbone upon unfolding of the various fingerprint domains.[29] Independent of the analysis method, however, accurate polymer elasticity models are required to quantify the hidden lengths of the folded proteins that are released by the unfolding events, giving rise to the limitations of PEG systems described above.

Adaptation of Surface Chemistry to Tether Protein Domains to ELP Linkers. The comparison of PEG with ELP linkers was carried out by cloning and recombinantly expressing two different ELPs both with 120 nm theoretical contour length ($ELP_{120\ nm}$, assuming 0.365 nm per amino acid).[30] One ELP linker contained an N-terminal sortase-tag ("GGG") and a C-terminal cysteine. The other ELP linker had a sortase-tag at its C-terminus ("LPETGG") and a cysteine at the N-terminus. Two analogous bioconjugation routes were used to attach ELP or PEG linkers to cantilevers and glass surfaces (Figure 2). To achieve a direct comparison, 15 kDa PEG linkers of similar contour lengths (~120 nm) were used. For PEG experiments, 15 kDa NHS-PEG-maleimide was immobilized onto an amino-silanized glass slide ($PEG_{120\ nm}$). The maleimide groups of the

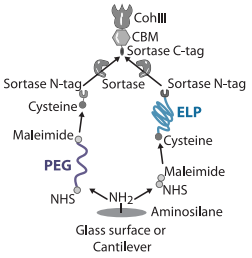


Figure 2. Comparison of immobilization strategies. For standard immobilization with PEG spacers, NHS chemistry was used to link PEG to amino-silanized surfaces. Protein constructs were then coupled *via* cysteine-sortase tag peptides to the maleimide end-groups on the PEG spacers. For immobilization with ELP linkers, a small-molecule NHS-maleimide cross-linker with a negligible contour length of 0.83 nm was used to couple cysteine-ELP spacers with a sortase-tag to the amino-silanized surface. In both cases, a fusion protein of interest, consisting of a CBM fingerprint domain and a mechanostable Coh receptor, was enzymatically coupled to the immobilized molecules on the surface in a subsequent step. Depicted is the functionalization of the glass surface with CohIII. The functionalization of the cantilever tip with DocIII followed a similar scheme.

PEG reacted with a GGGGG-Cys peptide, leaving the sortase N-tag available for subsequent derivatization. For ELP experiments, a small-molecule cross-linker (sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate, sulfo-SMCC), which added negligible contour length (0.83 nm) to the system, was first immobilized onto amino-silanized glass, followed by coupling with $GGG\text{-}ELP_{120\ nm}$-Cys. Both strategies resulted in the sortase N-tag being available for conjugation *via* sortase-mediated enzymatic ligation. The protein of interest (CohIII-CBM-LPETGG) was linked by sortase A to ELP or PEG (Figure 2). The same strategy was used for the cantilever, except GGG-Xmod-DocIII was conjugated by sortase A to $Cys\text{-}ELP_{120\ nm}$-LPETGG or to $PEG_{120\ nm}$-coupled Cys-LPETGG. Our enzyme-mediated protein immobilization approach has the advantage of site-specific linkages and results in a homogeneous

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346–6354

orientation of the proteins at the surface. Such uniformly immobilized proteins lead to a well-defined propagation of the applied force through the molecular complex under investigation and to well-defined distributions of the unfolding/rupture events in the force−extension curves. The use of N- and C-terminal tags for surface chemistry also ensured that only full-length (*i.e.*, fully translated) ELPs were measured in the experiment.

AFM experiments performed with ELPs as linkers showed a higher percentage of clearly identifiable single-molecule unfolding traces. We attribute this to the bulky character of the ELPs. They provide a less dense surface immobilization of the biomolecules of interest when compared to PEG-based immobilization. This behavior is advantageous since high surface density frequently causes multiple interactions between surface- and cantilever-bound molecules in SMFS experiments (Supplemental Figure S1). Multiple interactions are generated when more than one receptor−ligand interaction is formed in parallel. The complicated unfolding and unbinding traces that result from multiple bonds pulled in parallel are hardly interpretable and therefore discarded from the analysis (Supplemental Figure S2). Efficient passivation of glass surfaces against nonspecific adhesion of proteins requires a dense PEG surface layer, to prevent proteins from nonspecifically sticking to the glass surface. Approaches such as titrating functional (*i.e.*, maleimide end-groups) with nonfunctional (*i.e.*, CH_3 end-groups) PEG or changing the concentration of binding agents or proteins of interest can improve the process. In our experience, however, surface immobilization with ELP instead of PEG linkers leads to better passivation of the surface and a higher percentage of single-molecule traces without the need for any titration of functional and nonfunctional linkers.

Comparison of Dispersity between PEG and ELP Linkers. All unfolding traces were presorted by an automated analysis routine, selecting for single interactions that display two consecutive CBM unfolding events. Following the automated sorting, deletion of obviously erroneous curves (typically 10%) caused by, for example, baseline drift was performed manually.[7,29] PEG unfolding traces showed wildly varying initial extensions prior to the first CBM unfolding event. This is likely caused by the non-negligible polydispersity of PEG, as we did not observe multiple discrete populations with ELP experiments. The intrinsic monodispersity of ELP molecules is a clear advantage. Since they are produced recombinantly in *E. coli* with functional tags *in vivo*, only full-length protein sequences have the necessary terminal peptide tags that allow for surface immobilization. Additionally, ELPs were purified with inverse transition cycling (ITC), a method developed for ELP purification based on their reversible precipitation behavior. Possibly shorter ELPs are removed during the process, since their cloud point is higher than for $ELP_{120\ nm}$. Although the polydispersity of chemically synthesized PEGs (mass distribution ∼10−20 kDa) is sufficiently low for many applications, it leads to a noticeable impact in SMFS.

The influence of PEG polydispersity on the SMFS data is illustrated in Figure 3A, which shows SMFS traces recorded with both PEG and ELP linkers and also shows example traces of the shortest and largest extensions found in a typical type III Coh:Doc data set. Figure 3B shows a histogram of extension values at which the first CBM unfolding event occurred. For ELPs, the distribution shows one peak centered at an extension value that is expected based on the known ELP linker length. In the case of the PEG experiment, however, three distinct

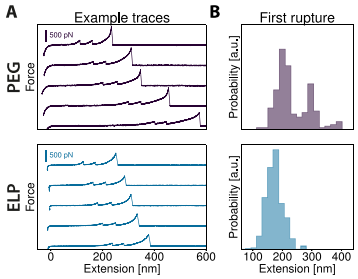


Figure 3. Comparison of dispersity of PEG and ELP linkers. (A) Typical force−extension traces for PEG (purple) and ELPs (blue). In the PEG linker experiment, the unfolding events occur over a wider range of absolute extension values, whereas unfolding events with ELP linkers occur over a narrow range. (B) Histograms showing the distribution of extension values corresponding to the first CBM unfolding event in each curve (PEG: *N* = 219; ELP: *N* = 521). Due to the polydispersity of the PEG linkers, three discrete populations with different extensions are clearly visible, while for ELPs only one population is observed.

populations are observed. This can be understood by considering that at the level of single molecules a polydisperse distribution results in discrete peaks representing the corresponding lengths of the discrete polymeric linkers on the cantilever tip. We interpret the distributions as being caused by three different PEG molecules with different lengths attached to the tip. Although the discrete distributions could conceivably be caused by different positions of the molecule attachment points to the AFM cantilever tip, this effect should be the same for ELPs. Moreover, varying linker lengths also reflect in varying steepness of the force−extension trace peaks, which would not occur simply because of attachment geometry (Figure 3A, PEG traces). We exclusively observed monomodal distributions for ELPs; therefore an anchor position effect seems not to play a major role. This polydispersity is clearly disadvantageous, since multiple linker lengths render data analysis more difficult. Curves cannot simply be overlaid in force−distance space due to varying loading rates. Furthermore, for constant-speed SMFS experiments, loading rate populations in dynamic force spectra will be broadened due to the probabilistic nature of the thermally driven rupture events.

We note that the PEG-modified surfaces are softer than ELP-modified surfaces during indentation of the tip into the polymer brush, as determined by the curvature at the beginning of each trace. The firmer ELP-modified surfaces require a lower indentation force to reach a linear force−distance regime after the initial soft indentation. For calibrating the inverse optical lever sensitivity, this is advantageous since high indentation forces can damage the molecules attached to the tip through adsorption and denaturation processes.[31]

Uniform ELP Stretching Behavior Minimizes Artifacts. We hypothesized that by replacing synthetic PEG linkers with biological ELP linkers, and thereby having a single type of polymer backbone throughout the mechanical system, better defined elasticity properties for the recording of force curves would be achievable. The persistence lengths of ELP peptide backbones should be comparable to those of unfolded protein

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

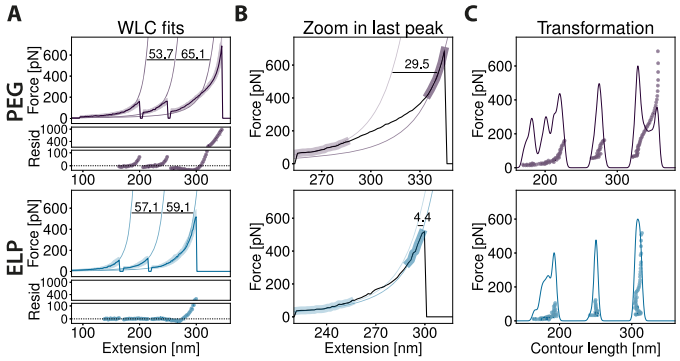


Figure 4. Elasticities of PEG and ELP linkers. (A) Superposition of multiple protein unfolding curves ("master curves") from SMFS experiments with PEG (purple, N = 73) and ELP linkers (blue, N = 151). The lower plots of each graph in panel A show the residuals of each WLC fit. Note that the residual plots are split into two subranges, shown in two windows from −35 to 120 pN (lower window) and from 120 to 1100 pN (upper window). The applied WLC model was extended by *ab initio* quantum mechanical calculations to correct for the enthalpic stretching of the polymer backbone.[33] Data were fitted with a fixed persistence length of 0.4 nm. The fits show that the stretching behavior of the mixed polymer system with PEG linkers deviates markedly at elevated forces from the predictions of the elasticity model, whereas the ELP curves agree reasonably well. (B) Final stretch and the Coh:Doc rupture event were fitted with the qmWLC model with two different contour lengths in the lower and upper force regime. The PEG molecules undergo a conformational transition,[16] resulting in different measured contour lengths for each force regime. For ELP molecules, a comparable transition was reported,[27,34] which apparently contributes to a much lower extent, so that SMFS experiments are much less affected. The differences in fitted contour length between the two fits are 29.5 nm for PEG linkers and 4.4 nm for ELP linkers. (C) Contour length transformations[29,35] of PEG and ELP master curves (purple and blue points). Ideally, the transformation results in data points aligning on vertical lines, where each line represents an energy barrier position for each stretching regime between two peaks in force−extension space. A KDE (Gaussian kernel, bandwidth: 2.5 nm) was calculated for the transformed data. The ELP data set showed the expected three peaks for the three unfolding and dissociation events, whereas the PEG data exhibit an irregular distribution with additional maxima.

domains, since they both consist of the same type of peptide-bonded polymer chains. This matching of the persistence length should be advantageous compared to PEG, which contains repeats of ethylene oxide groups with lower stiffness. Accurate description of the mechanical system under investigation by elasticity models plays a crucial role in determining characteristic parameters such as persistence lengths and contour length increments.

Previous studies had shown that at forces below 100 pN PEG elasticity may be satisfactorily described by standard elasticity models.[16] In a systematic study in this force range, we compared ELP and PEG linkers and corroborated these earlier results. The data and a thorough discussion thereof are given in the Supporting Information (see particularly Supplemental Figure S3).

At elevated forces, however, stretching of PEG through its conformational transition causes marked deviations from ideal polymer behavior. In aqueous environments, water molecules bridge neighboring ethylene oxide monomers by hydrogen bonding to two adjacent oxygen groups in the PEG backbone. By this means, water stabilizes the trans−trans−gauche configuration with a binding energy of around 3 kT. When PEG is stretched, however, the subunits of the backbone are forced increasingly into a slightly longer all-trans configuration and the bound water molecules are released. This conformational change, which contributes prominently to the polymer elasticity in the force range of up to ca. 300 pN, causes an increase in the measured net contour length of the polymer backbone.[16,17]

Figure 4A shows assemblies of multiple data traces ("master curves") of PEG- and ELP-linked proteins, respectively. The master curves are obtained by first aligning force−extension traces along the extension axis using an algorithm to maximize cross-correlation values in contour length space and then finding most probable force values of aligned traces in force distance space (see the Materials and Methods section). A recently introduced worm-like chain (WLC) approximation model[32] with an *ab initio* quantum mechanical correction for backbone stretching at high forces[33] (qmWLC) was then fitted to the traces with a fixed persistence length of 0.4 nm.

In the case of PEG linkers, a pronounced linear regime between 100 and 300 pN is visible in the last stretch prior to Coh:Doc rupture. As a consequence, the qmWLC cannot model this polymer correctly. ELPs do not show such a conformational change to this extent, and therefore the elasticity model fits satisfyingly. A fitting approach where the persistence length is also a free fit parameter is shown in Supplemental Figure S4. This approach misused the persistence length to compensate for the gauche-to-trans conformational change in the polymer; therefore, it resulted in largely unrealistic values for the contour length increments.

Figure 4B shows details of the last stretch before the Coh:Doc dissociation, highlighting the difference between PEG and ELP linkers. Two separate fits in the respective low- and high-force regimes illustrate the differences in polymer length before and after the conformational transition. We note that ELPs were also reported to have a force-induced conformational change, in this case based on proline cis−trans

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

isomerization that also extends the contour length.[27,34] However, the low number of prolines in the overall sequence (every fifth amino acid) in the ELP motif renders this effect much smaller compared to the conformational change of PEG and will be camouflaged by signal noise in typical experiments with proteins.

Figure 4C shows the transformation into contour length space using the qmWLC model. A kernel density estimate (KDE) was used (Gaussian kernel, bandwidth of 2.5 nm) to generate smooth functions describing the contour length increments observed between unfolding or rupture events, which in this case included 2× CBM unfolding and Coh:Doc dissociation. In the case of PEG linkers, the KDE−contour length distribution shows several peaks. This is because of the failure of the qmWLC model to accurately describe the force response of the polymer. Determining the contour length increments between the peaks of the KDE proves problematic even for this relatively simple exemplary case of two large fingerprint unfolding events and a receptor ligand dissociation. Smaller unfolding steps or even folding intermediates, which appear as substeps, would be even harder to pinpoint with the PEG system. In the case of ELP-immobilized proteins, only three distinct peaks appear, with much more clearly identifiable contour length increments between the peaks.

CONCLUSION

PEG linkers have successfully been employed in numerous studies to anchor biomolecules of interest to surfaces for SMFS. In the low-force regime (below 100 pN) the extended WLC model describes their elastic properties with sufficient accuracy for the majority of applications. For elevated forces, however, the conformational transitions in the PEG backbone would necessitate further development of elasticity models for a convincing description.[16] Moreover, the inherent polydispersity of PEGs, together with their complex elasticity, complicates data analysis and reduces the amount of information that can be deduced from SMFS.

The ELP-based linkers, however, have proven in our studies to be significantly improved linker molecules for surface immobilization and passivation purposes in single-molecule force experiments. ELPs are monodisperse, are highly flexible, and readily allow for direct, site-specific tethering. We showed that these features lead to more accurate measurements of contour length increments in receptor−ligand polyprotein force spectroscopy experiments. A well-established elasticity model suffices for the data analysis.

Even at low forces, the PEG subunits already start to change their conformational state occupancy. At 50 pN, the probability for their elongated state is already above 10%.[16] Therefore, the findings we present here are also relevant for investigations at lower forces or in systems that should be analyzed over a large range of forces. PEG linkers may still deliver satisfying results, as long as data in similar force ranges can be compared. In some cases, elasticity parameters such as the Kuhn length or persistence length can heuristically compensate for effects not explicitly described by the model. As soon as different force ranges of multiple domains need to be compared, though, the varying proportions of elongated (all-trans) *versus* non-elongated (trans−trans−gauche) PEG subunits cannot simply be accounted for by the elasticity parameter, and therefore measured contour length increments get distorted. Different biochemical approaches like those described here are thus necessary to gain meaningful insights. These scenarios include, for example, shielded unfolding events or small substeps, where the force cannot drop sufficiently in between stretching events.

The ELPs investigated here represent only one formulation of the vast variety of smart polymer linkers that could be utilized in SMFS experiments. Further studies are required to evaluate other nonstructured, non-proline-containing protein linkers to determine their suitability for SMFS studies, since the amino acid side chain composition may affect the persistence length[36,37] or give rise to nonentropic behavior. Biotechnological characteristics, *i.e.*, recombinant production yields and ease of purification, are as important as the biophysical requirements, which renders the easily produced ELPs particularly attractive. Other smart polymers should be similarly accessible to perform as suitable alternatives. The reported approach can be applied to enhance SMFS studies with purified proteins on functionalized surfaces as shown here or alternatively to modify cantilevers for chemical recognition imaging and force spectroscopy on artificial membranes or cell surfaces. It can easily be adopted by standard molecular biology equipped laboratories to streamline the procedure and improve data quality for resolving smaller unfolding features with high accuracy. Studies on smart polymers as tethers for SMFS experiments might also help to develop environmentally responsive surfaces, which bear potential for exciting applications in the nanobiosciences.

MATERIALS AND METHODS

All reagents were at least of analytical purity grade and were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Carl Roth GmbH (Karlsruhe, Germany). All buffers were filtered through a 0.2 μm poly(ether sulfone) membrane filter (Nalgene, Rochester, NY, USA) prior to use. The pH of all buffers was adjusted at room temperature.

A 300 amino acid long ELP was the basis for the AFM linker constructs used in this study, and the underlying cloning and protein purification procedure of the ELP is described in detail elsewhere.[19] The ELP sequence was $[(VPGVG)_5\text{-}(VPGAG)_2\text{-}(VPGGG)_3]_6$ and is referred to as $ELP_{120\ nm}$.

Standard molecular biology laboratories capable of producing recombinant proteins are equally capable of expressing ELPs, since both rely on the same principles, reagents, and instrumentation. With our plasmids provided at Addgene, cloning can even be avoided and production of ELP linkers for protein immobilization can be performed right away.

Cloning. A detailed description of the cloning procedure of the constructs can be found in the Supporting Information (Figures S5−S11). ELP sequences used in this study, along with 40 nm length variants and binding handles, are deposited at Addgene and available upon request (Addgene accession numbers: 90472: Cys-$ELP_{120\ nm}$-LPETGG, 90475: Cys-$ELP_{40\ nm}$-LPETGG, 91571: GGG-$ELP_{40\ nm}$-Cys, 91572: GGG-$ELP_{120\ nm}$-Cys, 91697: CohIII-CBM-HIS-LPETGG, 91698: GGG-HIS-CBM-Xmod-DocIII).

Transformation of Cells. A 2 μL amount of Gibson assembly or ligation reaction transformed *DH5α* cells (Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 1 min at 42 °C, 1 h at 37 °C in SOC medium) was used. The cells were plated on 50 μg/mL kanamycin-containing LB agar and incubated overnight at 37 °C. Clones were analyzed with Colony PCR, and clones with amplicons of appropriate lengths were sent to sequencing.

Protein Expression. Chemically competent *E. coli NiCo21(DE3)* (New England Biolabs, Ipswich, MA, USA) were transformed with 50 ng of plasmid DNA for the expression of all constructs used in this study. Transformed cells were incubated in autoinduction ZYM-5052 media (for ELP containing constructs supplemented with 5 mg/mL proline, valine, and 10 mg/mL glycine; 100 μg/mL kanamycin) for 24 h (6 h at 37 °C, 18 h at 25 °C).[38] Expression cultures were harvested *via* centrifugation (6500*g*, 15 min, 4 °C), the supernatant was discarded, and the pellets were stored at −80 °C until further lysis.

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

Throughout the whole purification process, for ELPs containing a cysteine, 1 mM tris(2-carboxyethyl)phosphine (TCEP, Thermo Fisher Scientific Inc., Waltham, MA, USA) or 1 mM of dithiothreitol (DTT) was added to the respective buffers. Cell pellets with proteins containing no HIS-tag were solubilized in 50 mM Tris-HCl pH 7.5 (supplemented with cOmplete, EDTA-free protease inhibitor cocktail, Sigma-Aldrich, St. Louis, MO, USA), and all other pellets in lysis buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 10% (w/v) glycerol, 0.1% (v/v) Triton X-100, 5 mM $MgCl_2$, DNase I 10 μg/mL, lysozyme 100 μg/mL).

Cys-$ELP_{120\ nm}$-LPETGG and GGG-$ELP_{120\ nm}$-Cys were purified with the ITC method.[39] After resolubilization, the cells were lysed by sonication (Bandelin Sonoplus GM 70, tip: Bandelin Sonoplus MS 73, Berlin, Germany; 40% power, 30% cycle, 2 × 10 min). The cells were kept on ice during the sonication procedure. The soluble fraction was separated from the insoluble cell debris by centrifugation (15000g, 4 °C, 1 h). In a first heating step (60 °C, 30 min) of the supernatant, most of the *E. coli* host proteins precipitated. The fraction of the collapsed ELPs was resolubilized by cooling the suspension for 2 h to 4 °C on a reaction tube roller. The insoluble host proteins were pelleted by centrifugation (15000g, 4 °C, 30 min). Further purification steps were necessary to increase the purity of the ELP solution. This was done by repeated thermoprecipitation of the ELP followed by redissolution.

The ELP solution was clouded by adding 1 M acetate buffer (final concentration 50 mM, pH 2.5) and 2 M NaCl. A heating step (60 °C, 30 min) ensured all ELPs were collapsed. A hot centrifugation (3220g, 40 °C, 75 min) was necessary to separate the high-salt, low-pH solution from the ELP pellet, which was resolubilized in 50 mM Tris-HCl (pH 7.0) after discarding the supernatant. The solution was incubated for 2 h at 4 °C to resolubilize all ELPs completely. A cold centrifugation step (3220g, 4 °C, 60 min) isolated the remaining insoluble fraction of the suspension. After decanting the supernatant, the salt concentration was increased and pH lowered, to precipitate the ELPs again. This cycle was repeated three times or extended if the purity of the solution was not high enough.

The constructs CohIII-CBM-HIS-LPETGG and GGG-HIS-CBM-Xmod-DocIII were expressed and lysed as described above. After the first centrifugation, the supernatant was, however, filtered (0.45 μm) and applied to a HisTrap FF (GE Healthcare Europe GmbH, Freiburg, Germany). Unspecifically bound proteins on the column were removed by washing five column volumes (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 20 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol). Finally, the desired HIS-tag containing protein was eluted (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 300 mM imidazole, Tween 20 0.25% (v/v), 10% (v/v) glycerol).

For long-term storage the protein solutions of the different constructs were concentrated (Amicon Ultra-15 centrifugal filter units 10K MWCO, Merck KGaA, Darmstadt, Germany) and reduced with 5 mM TCEP overnight (at 4 °C) for constructs that contained a cysteine. The buffer of the reduced ELP solution was exchanged (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.) to 50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, with a pH of 7.2, and 10% (v/v) glycerol and flash frozen in liquid nitrogen in small aliquots to be stored at −80 °C. All other proteins were exchanged with 25 mM Tris-HCl, 75 mM NaCl, and 5 mM $CaCl_2$ with a pH of 7.2 and supplemented with a final glycerol concentration of 20% (v/v). No loss of functionality of the ELPs (cross-linking and passivation capability) could be detected, when stored buffered or lyophilized in small aliquots at −80 °C, over the duration of more than one year.

SDS-PAGE (Any kD Mini-PROTEAN stain-free gels, Bio-Rad Laboratories GmbH, Hercules, CA, USA) was employed to detect any impurities. Since ELPs could not be stained with the stain-free technology, an Alexa Fluor 647-C_2-maleimide dye (Thermo Fisher Scientific Inc.) was incubated for 1 h at room temperature with the ELP solution. An appropriately diluted protein solution was mixed with 5× loading buffer (250 mM Tris-HCl, pH 8.0, 7.5% (w/v) SDS, 25% (v/v) glycerol, 0.25 mg/mL bromophenol blue, 12.5% (v/v) 2-mercaptoethanol) and heated for 5 min at 95 °C.

ELP concentration was photometrically determined at 205 nm (Ultrospec 3100 Pro, Amersham Biosciences, Amersham, England, and TrayCell, Hellma GmbH & Co. KG, Müllheim, Germany). For all other constructs an absorption measurement at 280 nm led to the concentration (NanoDrop UV−vis spectrophotometer, Thermo Fisher Scientific Inc.). The extinction coefficient was determined theoretically for ELPs at 205 nm[40] and 280 nm[41] for all other fusion proteins.

AFM Sample Preparation. Force spectroscopy samples, measurements, and data analysis were prepared and performed according to previously published protocols.[10,35] Silicon nitride cantilevers (Biolever mini, BL-AC40TS-C2, Olympus Corporation, Tokyo, Japan; nominal spring constant: 100 pN/nm; 25 kHz resonance frequency in water) were used as force probes. Surface chemistry for cantilevers was similar to that for coverslips (Menzel Gläser, Braunschweig, Germany; diameter 24 mm). Surfaces were amino-silanized with 3-(aminopropyl)dimethylethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany). α-Maleinimidohexanoic-ω-NHS PEG (NHS-PEG-Mal, Rapp Polymere, Tübingen, Germany; PEG-MW: 15 kDa) was used as a linker for the sortase peptides (GGGGG-C and C-LPETGG, Centic Biotec, Heidelberg, Germany) in PEG-linked experiments. The cysteine-containing ELPs were linked to the surface with a sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate) cross-linker (sulfo-SMCC, Thermo Fisher Scientific Inc.). PEG or cross-linker (10 mM) was dissolved in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5.

Sortase-catalyzed coupling of the fingerprint molecules (GGG-CBM-Xmod-DocIII and CohIII-CBM-LPETGG) was done in 25 mM Tris-HCl, pH 7.2, 5 mM $CaCl_2$, and 75 mM NaCl at 22 °C for 2 h. Typically, 50 μM ELP or sortase peptide was coupled with 25 μM fingerprint molecule and 2 μM sortase enzyme.

In between both of the cross-linking steps (PEG, SMCC, or ELP, peptide reaction) surfaces were rinsed with water and dried with nitrogen. After immobilization of the fingerprint molecules, surfaces were rinsed in measurement buffer (25 mM Tris-HCl, pH 7.2, 5 mM $CaCl_2$, 75 mM NaCl). The reaction of the different surface chemistry was done spatially separated by using silicone masks (CultureWell reusable gaskets, Grace Bio-Laboratories, Bend, OR, USA). The mask was applied after silanization and removed under buffer after the last immobilization step.

AFM-SMFS Measurements. Data were taken on custom-built instruments (MFP-3D AFM controller, Oxford Instruments Asylum Research, Inc., Santa Barbara, CA, USA; piezo nanopositioners: Physik Instrumente GmbH & Co. KG, Karlsruhe, Germany, or Attocube Systems AG, Munich, Germany).

Instrument control software was custom written in Igor Pro 6.3 (Wavemetrics Inc., Portland, OR, USA). Piezo position was controlled with a closed-loop feedback system running internally on the AFM controller field-programmable gate array. A typical AFM measurement took about 12 h and was done fully automated and at room temperature. Retraction velocity for constant-speed force spectroscopy measurements was 0.8 μm/s. Cantilever spring constants were calibrated after completing all measurements on different spots on the surface using the same cantilever. This was done by utilizing the thermal method applying the equipartition theorem to the one dimensionally oscillating lever.[31,42]

Force−Extension Data Analysis. Obtained data were analyzed with custom-written software in Python (Python Software Foundation, Python Language Reference, version 2.7, available at http://www.python.org), utilizing the libraries NumPy, SciPy, and Matplotlib.

Raw voltage data traces were transformed into force distance traces with their respective calibration values after determining the zero force value with the baseline position. A correction of the force-dependent cantilever tip *z*-position was carried out. Force distance traces were filtered for traces showing two CBM unfoldings and a subsequent type III cohesin−dockerin dissociation, without preceding Xmodule unfolding.[7] This screening was carried out by detecting maximum-to-maximum distances of kernel density estimate (Gaussian kernel, bandwidth 1 nm) peaks in contour length space in each single trace, after applying thresholds for force, distance, and number of peaks. For

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

sorting data sets, transformation of force distance data into contour length space was done with a manually fixed persistence length of 0.4 nm, to measure distances of energy barrier positions.[29,43] Sorting was done allowing generous errors to the expected increments to account for the conformational stretching of the spacer molecules. Fits to the force−extension data with the WLC model had the following parameters additionally to the values mentioned in the figure captions, if not stated otherwise: initial guess for persistence length: 0.4 nm; fit precision: 1×10^{-7}. For assessment of transformation quality, the inverse worm-like-chain model was applied for transformation of force distance traces into the contour length space in a force window of 10 to 125 pN and with a persistence length previously fitted to each peak separately: The global mean value of each data set for each peak was used. Final alignments of the whole data sets were assembled by cross-correlation.

Master Curve Assembly. The master curves were assembled by cross-correlation of each force−distance trace of a presorted data set with all previous curves in contour length space, starting with a random curve. Each curve was shifted on its x axis to fit the maximum correlation value and added to the set assembly in contour length space. Subsequently, a second run was performed, cross-correlating each curve with the previously assembled set, to facilitate an equal correlation template for every curve, independent of its occurrence. Finally, the most probable shift was calculated by a KDE and subtracted from each curve to get representative absolute distances with respect to the origin. Distance and correlation value thresholds were applied to filter out less probable PEG populations and otherwise badly fitting data. In a final step, all overlaid raw data points in force−distance space were binned on the x axis into nanometer-sized slices, and their densities on the y axis were estimated by a KDE for each slice. Near the rupture events, where the kernel density estimates cannot unambiguously identify maxima of the data slices, the value was set to zero. Therefore, after each rupture, a small "gap" is visible, which was not included in data points used for fitting. Their most probable value and the corresponding full width at half-maxima then assembled the master curve. Although by this procedure representative absolute rupture forces for the domains are not necessarily reproduced to the highest accuracy, the most probable and most representative pathway of the elastic behavior in between peaks is resembled well.

qmWLC model. For WLC fits and transformations into contour length space, a recently improved approximation, solved for the extension, was used,[32] adding correction terms for quantum mechanical backbone stretching.[33]

With the abbreviations

$$f = FL_p/kT \tag{1}$$

$$b = \exp\left(\sqrt[4]{\frac{900}{f}}\right) \tag{2}$$

WLC fits were done with the model formula

$$x = L_{corr}\left(\frac{4}{3} - \frac{4}{3\sqrt{f+1}} - \frac{10b}{\sqrt{f}(b-1)^2} + \frac{f^{1.62}}{3.55 + 3.8f^{2.2}}\right) \tag{3}$$

With the quantum mechanical correction,

$$L_{corr} = \frac{L_{c,0}}{2y_2}\left(\sqrt{4Fy_2 + y_1^2} - y_1 + 2y_2\right) \tag{4}$$

where y_1 and y_2 are the *ab initio* parameters from the original publication.

Transformations were performed with the model contour length:

$$L_c = \frac{x}{\frac{4}{3} - \frac{4}{3\sqrt{f+1}} - \frac{10b}{\sqrt{f}(b-1)^2} + \frac{f^{1.62}}{3.55 + 3.8f^{2.2}}} \tag{5}$$

With the reverse quantum mechanical correction for zero force contour length,

$$L_{c,0} = \frac{L_c}{\frac{1}{2y_2}\left(\sqrt{y_1^2 + 4y_2F} + 2y_2 - y_1\right)} \tag{6}$$

with x being the extension, L_c the model contour length, F the force, L_p the persistence length, k Boltzmann's constant, T the temperature, y_1 and y_2 the quantum mechanical correction parameters, L_{corr} the qm-corrected contour length, and $L_{c,0}$ the reverse qm-corrected contour length at zero force. As a nonlinear fitting algorithm, a Levenberg−Marquardt least-squares minimization method was applied.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsnano.7b02694.

Further details on experimental methods, supplementary results, and sequence information (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: gaub@lmu.de.

ORCID

Hermann E. Gaub: 0000-0002-4220-6088

Author Contributions

⊥W. Ott and M. A. Jobst contributed equally to this work.

Author Contributions

W.O.: experiment design, sample preparation, measurements, data analysis, writing of manuscript; M.A.J.: experiment design, data analysis, writing of manuscript; M.S.B.: data analysis; E.D.: sample preparation; L.F.M.: data analysis; M.A.N.: experiment design, writing of manuscript; H.E.G.: experiment design, writing of manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Advanced Grant "Cellufuel" of the European Research Council and the Deutsche Forschungsgemeinschaft through SFB 1032. M.A.N. acknowledges support from an ERC Starting Grant "Molecular Mechanical Adhesives" number 715207 and from Society in Science−the Branco Weiss Fellowship from ETH Zurich. We thank T. Verdorfer and C. Schoeler for proofreading and helpful discussions.

REFERENCES

(1) Cao, Y.; Li, H. Engineered Elastomeric Proteins with Dual Elasticity Can Be Controlled by a Molecular Regulator. *Nat. Nanotechnol.* **2008**, *3*, 512−516.

(2) Lv, S.; Dudek, D. M.; Cao, Y.; Balamurali, M. M.; Gosline, J.; Li, H. Designed Biomaterials to Mimic the Mechanical Properties of Muscles. *Nature* **2010**, *465*, 69−73.

(3) Rivas-Pardo, J. A.; Eckels, E. C.; Popa, I.; Kosuri, P.; Linke, W. A.; Fernández, J. M. Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep.* **2016**, *14*, 1339−1347.

(4) Ott, W.; Jobst, M. A.; Schoeler, C.; Gaub, H. E.; Nash, M. A. Single-Molecule Force Spectroscopy on Polyproteins and Receptor−ligand Complexes: The Current Toolbox. *J. Struct. Biol.* **2017**, *197*, 3−12.

(5) Bull, M. S.; Sullan, R. M. A.; Li, H.; Perkins, T. T. Improved Single Molecule Force Spectroscopy Using Micromachined Cantilevers. *ACS Nano* **2014**, *8*, 4984−4995.

(6) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. Single-Molecule Dissection of the High-

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

Affinity Cohesin-Dockerin Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.
(7) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; E, G. H.; Nash, M. A. Ultrastable Cellulosome-Adhesion Complex Tightens under Load. *Nat. Commun.* **2014**, *5*, 1−8.
(8) Baumann, F.; Bauer, M. S.; Milles, L. F.; Alexandrovich, A.; Gaub, H. E.; Pippig, D. A. Monovalent Strep-Tactin for Strong and Site-Specific Tethering in Nanospectroscopy. *Nat. Nanotechnol.* **2015**, *11*, 89−94.
(9) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen Clostridium Perfringens. *J. Phys. Chem. B* **2017**, *121*, 3620−3625.
(10) Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K. Thiol-Based, Site-Specific and Covalent Immobilization of Biomolecules for Single-Molecule Experiments. *Nat. Protoc.* **2010**, *5*, 975−985.
(11) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E690−E697.
(12) Popa, I.; Rivas-Pardo, J. A.; Eckels, E. C.; Echelman, D.; Valle-Orero, J.; Fernandez, J. M. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J. Am. Chem. Soc.* **2016**, *138*, 10546−10553.
(13) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(14) Pippig, D. A.; Baumann, F.; Strackharn, M.; Aschenbrenner, D.; Gaub, H. E. Protein-DNA Chimeras for Nano Assembly. *ACS Nano* **2014**, *8*, 6551−6555.
(15) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. From Genes to Protein Mechanics on a Chip. *Nat. Methods* **2014**, *11*, 1127−1130.
(16) Oesterhelt, F.; Rief, M.; Gaub, H. E. Single Molecule Force Spectroscopy by AFM Indicates Helical Structure of Poly(ethylene-Glycol) in Water. *New J. Phys.* **1999**, *1*, 1−11.
(17) Liese, S.; Gensler, M.; Krysiak, S.; Schwarzl, R.; Achazi, A.; Paulus, B.; Hugel, T.; Rabe, J. P.; Netz, R. R. Hydration Effects Turn a Highly Stretched Polymer from an Entropic into an Energetic Spring. *ACS Nano* **2017**, *11*, 702−712.
(18) Xue, Y.; Li, X.; Li, H.; Zhang, W. Quantifying Thiol-Gold Interactions towards the Efficient Strength Control. *Nat. Commun.* **2014**, *5*, 4348.
(19) Ott, W.; Nicolaus, T.; Gaub, H. E.; Nash, M. A. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. *Biomacromolecules* **2016**, *17*, 1330−1338.
(20) Tang, N. C.; Chilkoti, A. Combinatorial Codon Scrambling Enables Scalable Gene Synthesis and Amplification of Repetitive Proteins. *Nat. Mater.* **2016**, *15*, 419−424.
(21) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules* **2010**, *11*, 944−952.
(22) Meyer, D. E.; Chilkoti, A. Genetically Encoded Synthesis of Protein-Based Polymers with Precisely Specified Molecular Weight and Sequence by Recursive Directional Ligation: Examples from the Elastin-like Polypeptide System. *Biomacromolecules* **2002**, *3*, 357−367.
(23) Gray, W. R.; Sandberg, L. B.; Foster, J. A. Molecular Model for Elastin Structure and Function. *Nature* **1973**, *246*, 461−466.
(24) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589*, 2477−2486.
(25) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the Specificity of Sortase Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13343−13348.
(26) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. Elastin: A Representative Ideal Protein Elastomer. *Philos. Trans. R. Soc., B* **2002**, *357*, 169−184.
(27) Valiaev, A.; Lim, D. W.; Oas, T. G.; Chilkoti, A.; Zauscher, S. Force-Induced Prolyl Cis-Trans Isomerization in Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2007**, *129*, 6491−6497.
(28) Valiaev, A.; Dong, W. L.; Schmidler, S.; Clark, R. L.; Chilkoti, A.; Zauscher, S. Hydration and Conformational Mechanics of Single, End-Tethered Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2008**, *130*, 10939−10946.
(29) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. Comparing Proteins by Their Unfolding Pattern. *Biophys. J.* **2008**, *95*, 426−434.
(30) Dietz, H.; Rief, M. Exploring the Energy Landscape of GFP by Single-Molecule Mechanical Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16192−16197.
(31) Proksch, R.; Schäffer, T. E.; Cleveland, J. P.; Callahan, R. C.; Viani, M. B. Finite Optical Spot Size and Position Corrections in Thermal Spring Constant Calibration. *Nanotechnology* **2004**, *15*, 1344−1350.
(32) Petrosyan, R. Improved Approximations for Some Polymer Extension Models. *Rheol. Acta* **2017**, *56*, 21−26.
(33) Hugel, T.; Rief, M.; Seitz, M.; Gaub, H. E.; Netz, R. R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments versus *Ab-Initio* Theory. *Phys. Rev. Lett.* **2005**, *94*, 048301.
(34) Valiaev, A.; Lim, D. W.; Schmidler, S.; Clark, R. L.; Chilkoti, A.; Zauscher, S. Hydration and Conformational Mechanics of Single, End-Tethered Elastin-like Polypeptides. *J. Am. Chem. Soc.* **2008**, *130*, 10939−10946.
(35) Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. Investigating Receptor-Ligand Systems of the Cellulosome with AFM-Based Single-Molecule Force Spectroscopy. *J. Visualized Exp.* **2013**, e50950.
(36) Stirnemann, G.; Giganti, D.; Fernandez, J. M.; Berne, B. J. Elasticity, Structure, and Relaxation of Extended Proteins under Force. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 3847−3852.
(37) Cheng, S.; Cetinkaya, M.; Gräter, F. How Sequence Determines Elasticity of Disordered Proteins. *Biophys. J.* **2010**, *99*, 3863−3869.
(38) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expression Purif.* **2005**, *41*, 207−234.
(39) MacEwan, S. R.; Hassouneh, W.; Chilkoti, A. Non-Chromatographic Purification of Recombinant Elastin-like Polypeptides and Their Fusions with Peptides and Proteins from Escherichia coli. *J. Visualized Exp.* **2014**, e51583.
(40) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci.* **2013**, *22*, 851−858.
(41) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.; Appel, R.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protocols Handbook* **2005**, 571−607.
(42) Hutter, J. L.; Bechhoefer, J. Calibration of Atomic-Force Microscope Tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868−1873.
(43) Jobst, M. A.; Milles, L. F.; Schoeler, C.; Ott, W.; Fried, D. B.; Bayer, E. A.; Gaub, H. E.; Nash, M. A. Resolving Dual Binding Conformations of Cellulosome Cohesin- Dockerin Complexes Using Single-Molecule Force Spectroscopy. *eLife* **2015**.

DOI: 10.1021/acsnano.7b02694
ACS Nano 2017, 11, 6346−6354

Supporting Information: Elastin-Like Polypeptide Linkers for Single Molecule Force Spectroscopy

Wolfgang Ott[a,b,⊥], Markus A. Jobst[a,⊥], Magnus S. Bauer[a], Ellis Durner[a], Lukas F. Milles[a], Michael A. Nash[c,d], Hermann E. Gaub[a,#]

[a] Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany.

[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 81377 Munich, Germany.

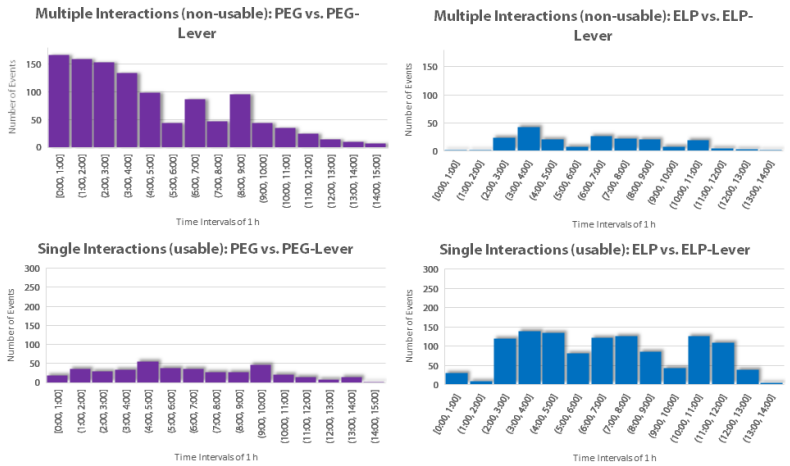
[c] Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

[d] Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland.

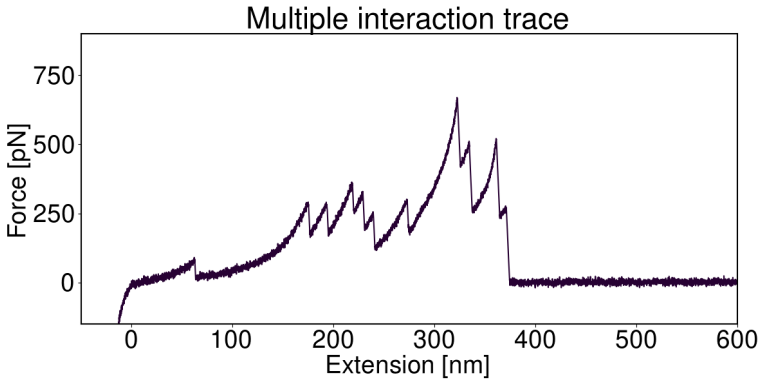
[⊥] These authors contributed equally to this work

[#] Corresponding author: gaub@lmu.de

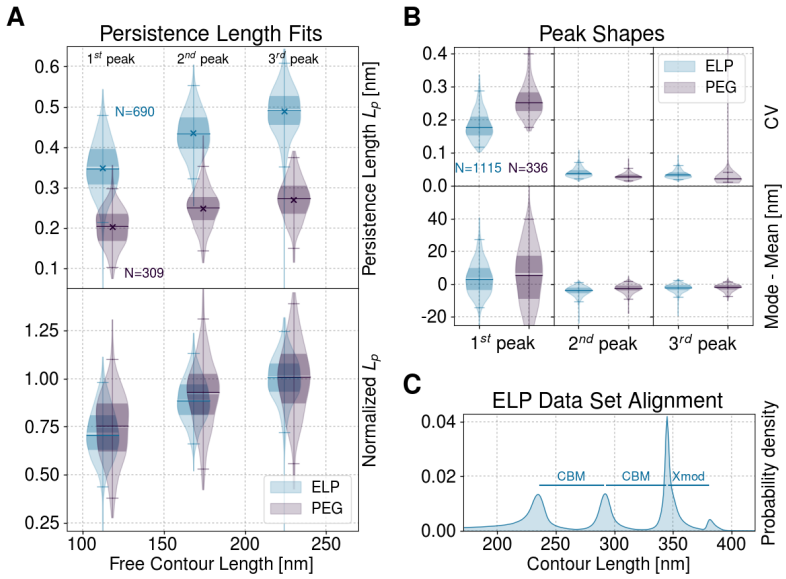
Supporting Information



Supplemental Figure S1. Number of curves within a 1 h timeframe were binned in one histogram bar. Multiple traces were traces with more than 10 peaks (**Supplemental Figure S2** shows an exemplary multiple interaction trace). Left (purple) is the PEG-lever versus the PEG-immobilization and right (blue) ELP-lever versus ELP-immobilization. The two top panels show number of multiple interactions over time. The bottom panels show number of single specific interactions over time.



Supplemental Figure S2. A typical example trace displaying multiple interactions.

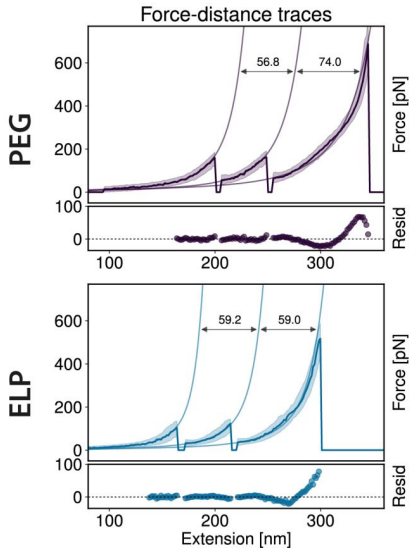


Supplemental Figure S3: Performance of contour length transformations. (A) Observed persistence lengths. Upper plot: observed persistence lengths preceding each CBM and Coh:DocIII unfolding/rupture peak as measured by WLC fits in the force range of 30 to 125 pN (ELP: 0.35, 0.44, and 0.49 nm; PEG: 0.20, 0.25, and 0.27 nm). Lower plot: same data normalized to the respective last peak means. The qualitative behavior over the unfolding of the peaks is similar for both constructs. **(B)** Assessment of transformation quality. Coefficient of variation (CV) as a measure of distribution broadness and distance of mode to mean as a measure of peak symmetry show better performance for ELP data for the first peaks. Later peaks show better performance of PEG data, although the differences are negligible. Transformations were done with the inverse WLC model only for data points between 10 and 125 pN. Persistence lengths for the transformations were chosen as the mean values of the WLC fits to each peak as shown in panel (A). **(C)** Alignment of transformed ELP curves in contour length space. Two CBM increments and one Xmod unfolding prior to Coh:Doc rupture are clearly detectable.

Low force performance of ELP linkers

For this analysis, only forces in a range from 10 to 125 pN were taken into account, to minimize the effects of conformational stretching. The elastic properties of the first stretching event of a data trace are dominated by the linker molecules. As more protein domains unfold, the peptide backbone of the unfolded domains contributes increasingly to the overall elastic response. Contour length transformations of force distance data were performed with the mean fitted persistence lengths of each peak, as shown in **Supplemental Figure S3, Panel A** (0.35, 0.44, and 0.49 nm for ELP data peaks; 0.20, 0.25, and 0.27 nm for PEG data peaks), to account for

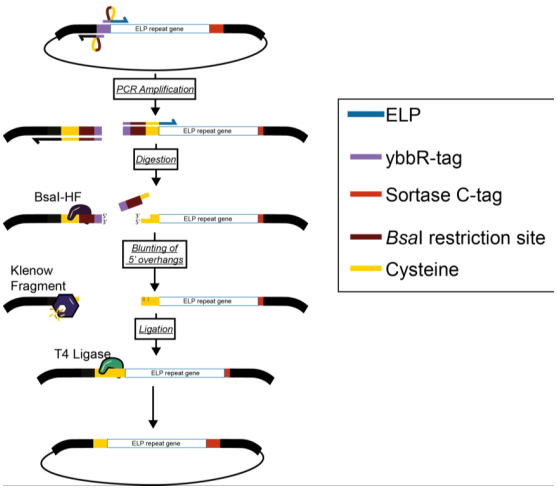
varying persistence lengths over the course of each pulling cycle. The persistence length as a measure for the stiffness of a polymer is lower for PEG than for ELP with bulky side chains and rotational restrictions of the peptide backbone. Comparable changes of persistence lengths over the course of an unfolding experiment were also observed earlier in other studies.[1,2] The distribution width and asymmetry of each peak in contour length space were evaluated separately by the coefficient of variation and the calculated difference of statistical mode and mean. A comparison of all datasets revealed that for the first unfolding peak, ELP datasets display slightly superior properties: the first peak for data with ELP linker tethering is sharper and more symmetric (**Supplemental Figure S3, Panel B**) as indicated by the narrower distribution and lower coefficient of variation. For the subsequent peaks 2 and 3, both PEG and ELP linkers perform similarly and the differences become negligibly small. Although the impact on data quality in this low force regime examined here, was not as severe as expected, ELP linkers seem to exhibit advantageous behavior for the first stretching events of each curve, and might improve accuracy in determining the following contour length increments to identify protein domains.



Supplemental Figure S4: Master curves fits with persistence lengths as an additional free fit parameter. If the persistence length is not kept fix, but also fitted to the data, it is clearly visible, that this parameter is optimized to compensate the conformational stretching effect for PEG datasets. While the qmWLC model fit itself looks better and has lower residuals compared to the fixed persistence length fit, the resulting contour length increment is way off and does not yield any meaningful value, rendering the model useless to extract information from the data. The two CBM domains have the exact same amino acid sequence and therefore should show the same contour length increments upon unfolding.

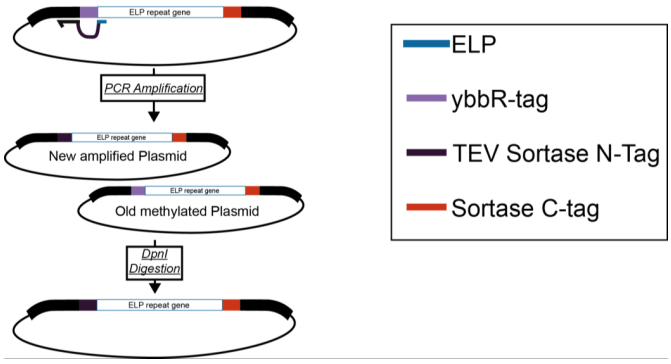
Cloning of ELP linkers. Standard PCR was used for amplification of DNA (Phusion High-Fidelity PCR Master Mix, Thermo Fisher Scientific Inc., Waltham, MA, USA). Melting temperatures were adjusted according to the employed primers (see **Table S1**, below).

A plasmid encoding ybbR-ELP$_{120\ nm}$-LPETGG described earlier[3] was modified to yield the plasmid Cys-ELP$_{120\ nm}$-LPETGG. PCR amplification of the plasmid with primers annealing at and downstream of the ybbR-tag was the first step (**Supplemental Figure S5**). The gene for the ELP is a highly repetitive sequence, hence it was necessary to anneal the forward primer at the ybbR-tag to create a unique attachment site. Since the ybbR-tag had to be removed, a *Bsa*I restriction site was incorporated with a primer downstream of the annealing region of the forward primer. The reverse primer had a cysteine encoded at its 5' end. After successful PCR amplification, the product was digested (*Bsa*I and *Dpn*I) and blunted (1h, 37°C, 5 Min, 80°C). The blunting reaction was performed in parallel with 1 µl of Klenow Fragment enzyme and the addition of 1 mM dNTPs (Thermo Fisher Scientific Inc., Waltham, MA, USA)).
After purification (QIAquick PCR purification kit or gel extraction kit (Qiagen GmbH, Hilden, Germany) the ligation reaction was set up: 1 µl of a T4 Ligase (10U/µl, Thermo Fisher Scientific Inc., Waltham, MA, USA was supplemented with 1 µl ATP (10 mM), 0.5 µl PEG-6000, 1 µl T4 Polynucleotide Kinase (PNK) and buffered in CutSmart buffer (New England Biolabs, Ipswich, MA, USA).



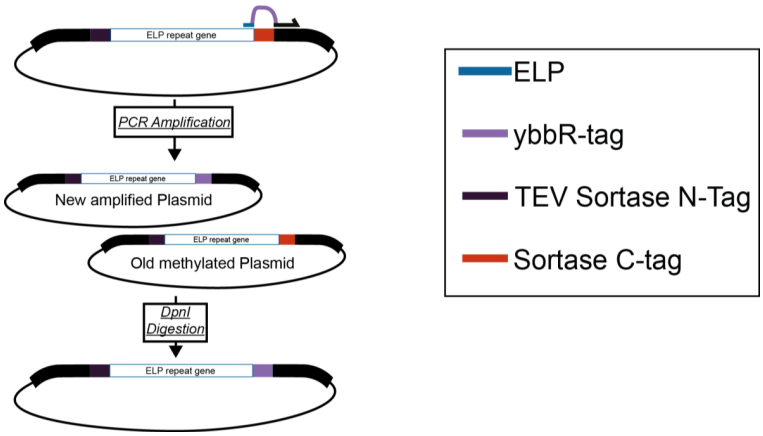
Supplemental Figure S5. Cloning scheme for Cys-ELP$_{120\ nm}$-LPETGG.

For the creation of the TEV-GGG-ELP$_{60\ nm}$-LPETGG plasmid, a plasmid encoding ybbR-ELP$_{60\ nm}$-LPETGG[1] was mutated with one QuikChange primer[4], annealing up- and downstream of the ybbR-tag introducing DNA encoding a TEV-site and a triple glycine. The TEV cleavage site was introduced to ensure full cleavage of the N-terminal methionine. This was assumed to be necessary, since Sortase A only works with glycines at the very N-terminal start of a protein. The QuikChange reaction was done with 50 ng DNA template, 1 µl of primer (10 pmol/µl) in 20 µl Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA, see **Supplemental Figure S6**).



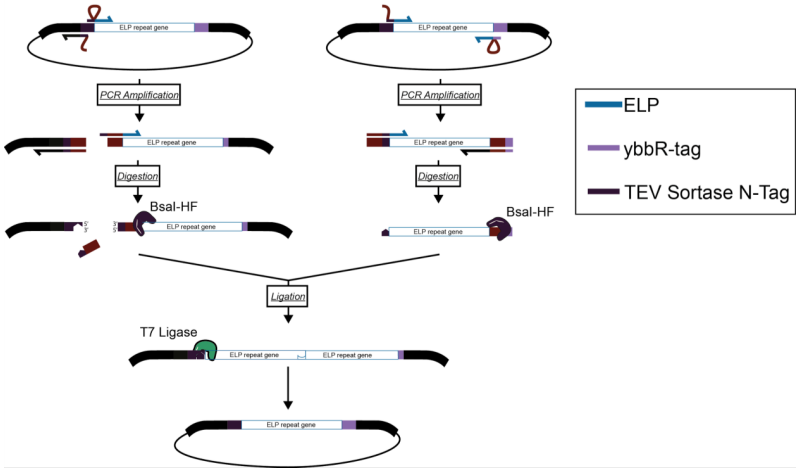
Supplemental Figure S6. Cloning scheme for TEV-GGG-ELP$_{60\ nm}$-LPETGG.

The newly obtained plasmid was modified again with QuikChange to exchange the C-terminal Sortase-tag with a ybbR-tag (**Supplemental Figure S7**).



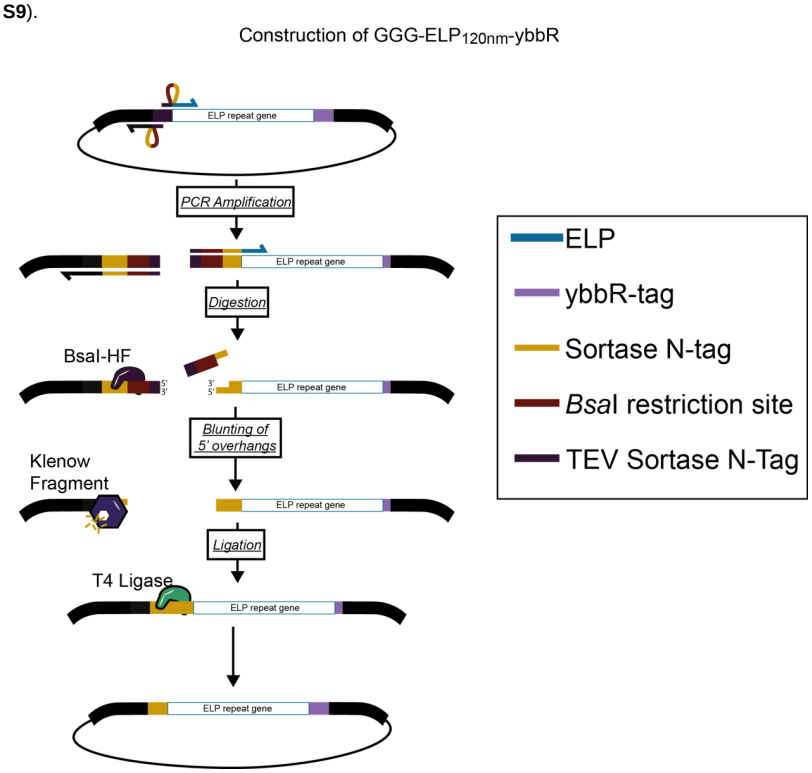
Supplemental Figure S7. Cloning scheme for TEV-GGG-$ELP_{60\ nm}$-ybbR.

The ELP gene cassette was duplicated by insertion of a gene sequence encoding $[(VPGVG)_5$-$(VPGAG)_2$-$(VPGGG)_3]_3$ into the linearized vector containing TEV-GGG-$ELP_{60\ nm}$-ybbR. This was done by GoldenGate cloning.[5] For this purpose, both vector and insert were amplified with primers encoding flanking *Bsa*I restriction sites. The *Bsa*I sites were designed to match the corresponding end of insert and backbone, without leaving any cloning scars. After *Bsa*I digestion and purification of the PCR product *via* gel extraction, both of the parts were ligated with their corresponding sticky ends (2.5 µl CutSmart buffer, 1.25 µl T7 ligase, 2.5 µl ATP (10 mM); New England Biolabs, Ipswich, MA, USA, see **Supplemental Figure S8**).



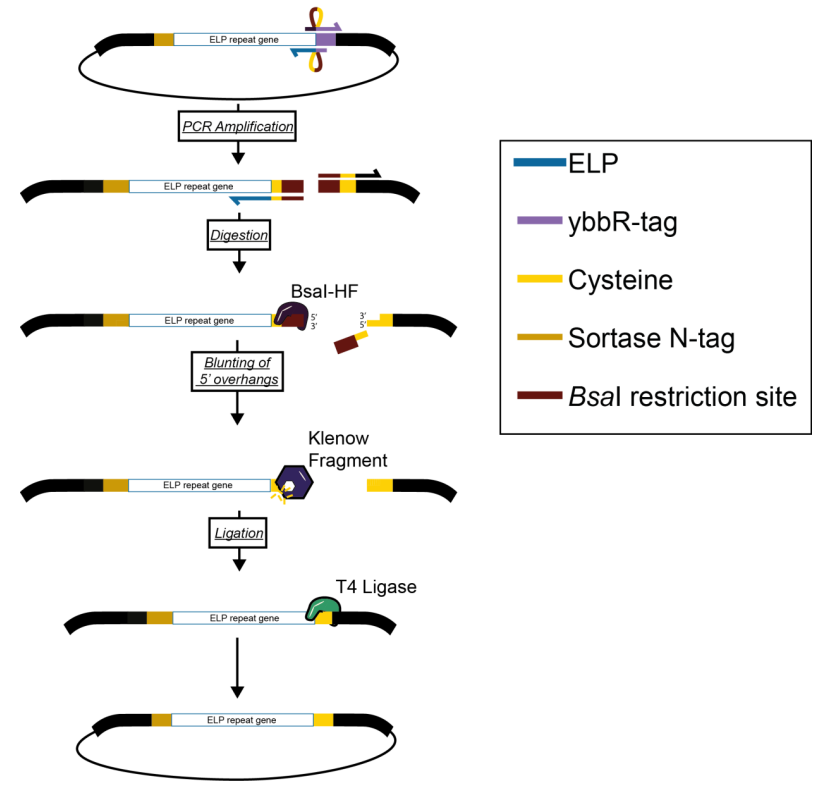
Supplemental Figure S8. Cloning scheme for TEV-GGG-$ELP_{120\ nm}$-ybbR.

Experiments showed that the *E. coli* methionine aminopeptidases already fully digested the N-terminal methionine proceeding the polyglycine. Hence, removal of the TEV cleavage site was desired to simplify the ELP production process. This was achieved by a linearization reaction, *Bsa*I digestion and religation as described above. Primers were designed to anneal at the TEV-site and encoded a *Bsa*I restriction site upstream of the triple glycine (**Supplemental Figure S9**).



Supplemental Figure S9. Cloning scheme for GGG-ELP$_{120\ nm}$-ybbR

Finally, the C-terminal ybbR-tag was switched to a cysteine. The reverse primer attached at the codons of the ybbR-tag with a *Bsa*I restriction site. The forward primer encoded a cysteine at its 5' end and annealed downstream of the stop codon. The linear plasmid was processed as described above (**Supplemental Figure S10**).

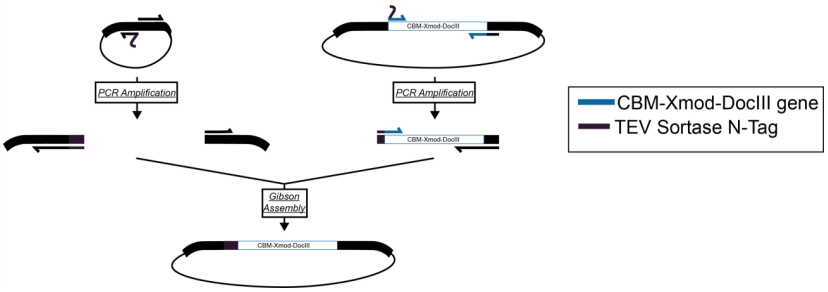


Supplemental Figure S10. Cloning scheme for GGG-ELP$_{120\ nm}$-Cys

Cloning of GGG-HIS-CBM-Xmod-DocIII and CohIII-CBM-HIS-LPETGG.
Basis for the construction were two plasmids published by Schoeler *et al.*[6] The plasmid encoding the gene for CohIII-CBM was linearized with primers encoding the Sortase C-tag. 4.5 µl of the PCR product was directly digested with 1 µl *Dpn*I (Thermo Fisher Scientific Inc., Waltham, MA, USA), 3' ends were phosphorylated with 1 µl T4 PNK (New England Biolabs, Ipswich, MA, USA) and the ends were religated with 1 µl T4 Ligase (10U/µl, Thermo Fisher Scientific Inc., Waltham, MA, USA) (15 Min at 37°C, 45 Min 22°C). The 10 µl reaction was supplemented with 1 µl ATP (10 mM), 0.5 µl PEG-6000 and 1 µl CutSmart buffer (10x, New England Biolabs, Ipswich, MA, USA).

The plasmid encoding the CohIII domain had a cloning scar (encoding the amino acids "GT") at the N-terminus. Glycine and threonine were removed since one single glycine is already reactive with the "LPETGG" in a Sortase A catalyzed reaction. This was done with a sequential linearization and religation reaction (as described above).

The CBM-Xmod-DocIII gene was subcloned with Gibson Assembly into a linearized vector with a TEV site followed by a Sortase N-tag. 10 µl of the HiFi MasterMix (2x, New England Biolabs, Ipswich, MA, USA), were mixed with a 10-fold molar excess of insert to the backbone (reaction volume 20 µl, 1 hr, 50°C; **Supplemental Figure S11**). Similar to the GGG-ELP$_{120\ nm}$-Cys, the unnecessary TEV site was removed, since *E. coli* already digested the N-terminal methionine sufficiently. This was achieved by employing the same procedure as described for CohIII-CBM linearization and religation.



Supplemental Figure S11. Cloning scheme for TEV-GGG-CBM-Xmod-DocIII

Supplemental Table S1. Overview of primers

Name	**Sequence (5'-3')**
Construction of Cys-ELP$_{120\ nm}$-LPETGG	
FW N-Cys BsaI	GACTCTCTGGAATTCATCGCTTCTAAACTGGC TGGTCTCCTGCGTGCCGGGAGAAGGAG
REV BsaI ybbR	CCCGGCACAGCCAGTTTAGAAGCGATGAATTC CAGAGAGTCGGTCTCACATATGTATATC
Construction of TEV-GGG-ELP$_{60\ nm}$-LPETGG	
QuikChange Primer ybbR to TEV-GGG	GACACCAGGGACTCCTTCTCCCGGCACACCG CCCCCTCCCTGGAAGTACAGGTTTTCCATATG TATATCTCCTTC
Construction of TEV-GGG-ELP$_{60\ nm}$-ybbR	
QuikChange Primer LPETGG to ybbR	GACACCAGGGACTCCTTCTCCCGGCACACCG CCCCCTCCCTGGAAGTACAGGTTTTCCATATG TATATCTCCTTC
Construction of TEV-GGG-ELP$_{120\ nm}$-ybbR	
FW backbone BsaI	GAAAACCTGTACTTCCAGGGAGGGGGGTCTC GGGGTGTGCCGGGAGAAGGAG
REV backbone BsaI	ATATATGGTCTCGACCGCCCCCTCCCTGGAAG TACAGGTTTTC
FW insert TEV-GGG BsaI	CCAGGGAGGGGGGTCTCGCGGTGTGCCGGG AGAAGGAG
REV insert BsaI	TCGAGTTAAGCCAGTTTAGAAGCGATGAATTC CAGAGAGTCGGTCTCCACCCTCACCCGG
Construction of GGG-ELP$_{120\ nm}$-ybbR	
FW ELP GGG	GGGGGCGGTGTGCCGGGAG
REV BsaI TEV	GGCACACCGCCCCCTCCCTGGAAGTACAGGT TTTCGGTCTCACATATGTATATCTCCTTC

Construction of GGG-ELP$_{120\ nm}$-Cys	
FW backbone Cys	GCCAGTTTAGAAGCGATGAATTCCAGAGAGTC GGTCTCCACCTTCACCC
REV ybbR BsaI	TGCTAACTCGAGTAAGATCCGGCTGCTAACAA AGCCC
Construction of GT-CohIII-CBM-HIS-LPETGG	
FW backbone	TAACTCGAGTAAGATCCGGCTGC
REV CBM LPETGG	GCCGCCGGTTTCCGGCAGCGGACCCTGGAAC AGAAC
Construction of CohIII-CBM-HIS-LPETGG	
FW CohIII	GCGCTCACAGACAGAGGAATG
REV backbone without GT	CATATGTATATCTCCTTCTTAAAGTTAA
Construction of TEV-GGG-HIS-CBM-XDocIII	
FW backbone	CTCGAGTAAGATCCGGCTGC
REV backbone	ACCGGGTTCTTTACCCC
FW insert	GTATGGGGTAAAGAACCCGGTGGCAGTGTAG TACCATC
REV insert	CGGATCTTACTCGAGTTATTCTTCTTCAGCATC GCCTG
Construction of GGG-HIS-CBM-XDocIII	
FW CBM	ATGGCCAATACACCGGTATCA
REV backbone	TCCGTGGTGGTGGTGGTGACCGCCCCCC ATATGTATATCTC

Supplemental Table S2. Biophysical parameters of the employed ELPs.

ELP repeats $(5)_x$	ε_{205} [1/M cm][7]	Molecular weight [Da][8]	Isoelectric point	Amino acids in ELP repeats (total)[8]	Total Length [nm][9] (.365 nm per aa)
Cys-ELP$_{120\ nm}$-LPETGG	851370	24763.08	3.20	300 (307)	112.06
GGG-ELP$_{120\ nm}$-Cys	843030	24379.63	3.23	300 (304)	110.96

Protein Sequences

GGG-ELP$_{120\ nm}$-Cys

Sortase N-Tag
ELP
Cysteine

GGGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGV
PGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVG
VPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGV
GVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPG
AGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVP
GGGVPGEGC

Cys-ELP$_{120\ nm}$-LPETGG
Cysteine
ELP
Sortase C-Tag

MCVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVP
GVGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGV
PGVGVPGVGVPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVG
VPGAGVPGAGVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGA
GVPGGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGAGVPGAGVPGGGVPG
GGVPGEGLPETGG

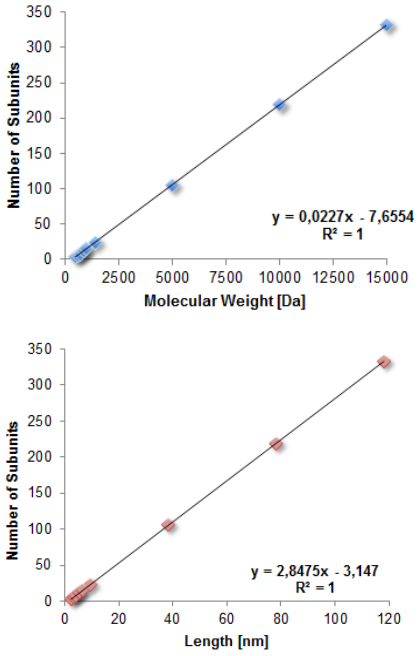
MGGG-HIS-CBM-Xmod-Dockerin III
Sortase N-Tag
His_6-Tag
CBM
Linker
Xmod
Dockerin III

MGGGHHHHHHGMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYT
VDGQKDQTFWSDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLE
PGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVPST
QPVTTPPATTKPPATTIPPSDDPNAVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTA
QIKKAVLHTVYNEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYE
IPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATT
VALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYTK
SSVDQYKDMAANEPNKLWDIVTGDAEEE

Cohesin III-CBM-HIS-LPETGG
Cohesin III
Linker
CBM
His_6-Tag
Sortase C-Tag

MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATT
GYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTV
ELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYL
VKANATYADGYIAIKAGEPGSVVPSTQPVTTPPATTKPPATTIPPSDDPNAMANTPVSGNLKVE
FYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNGI
TSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKND
WSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGELKLPRSRHHHHHHGSLEVLFQ
GPLPETGG

Linker Length. The artefacts generated by PEG linkers at elevated forces can be reduced by shortening the linker molecules. Usually our force spectroscopy experiments employ spacers with 40 nm length. Many SMFS assays utilize these 5 kDa PEG linkers, where the effect is scaled down proportionally with length, however still present. Further truncation would minimize the influence of the conformational change of PEG spacers, but in return raise other concerns: i) reduced mechanical isolation of the molecules under investigation by low pass filtering from transducer oscillations, to ensure purely thermally driven unfolding and dissociation events and defined loading rates[10], ii) reduced passivation of the surfaces against nonspecific adsorption, and iii) influence of surface effects and effects of the linker molecules themselves on the domains of interest. Employing peptide based smart polymers as linkers offer a new solution to this issue, avoiding linker artefacts almost entirely.



Supplemental Figure S12. Conversion of PEG molecular weights with functional end groups into their corresponding lengths. Based on the molecular weight of PEGs with functional groups maleimide and NHS, the number of subunits for various PEGs can be determined. Subsequently, the PEG contour lengths for a given number of subunits can be calculated. The data were obtained from the NHS-PEG-maleimide portfolio of Thermo Scientific and Rapp Biopolymers.

Supplemental Table S3. Overview of average molecular weight and length of PEG-Polymers. In blue are the calculated polymer sizes, in black the data the calculation is based on. Number of subunits were always round to the next integer.

Molecular Weight [Da]	Number of Subunits	Length [nm]
513.3	4	2.5
601.6	6	3.2
689.71	8	3.9
865.92	12	5.3
1394.55	24	9.5
1000	15	6.4
5000	106	38.3
10000	220	78.1
15000	333	118.0

References

(1) Liu, R.; Garcia-Manyes, S.; Sarkar, A.; Badilla, C. L.; Fernández, J. M. Mechanical Characterization of Protein L in the Low-Force Regime by Electromagnetic Tweezers/evanescent Nanometry. Biophys. J. 2009, 96, 3810–3821.
(2) Walther, K. A.; Gräter, F.; Dougan, L.; Badilla, C. L.; Berne, B. J.; Fernandez, J. M. Signatures of Hydrophobic Collapse in Extended Proteins Captured with Force Spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2007, 104, 7916–7921.
(3) Ott, W.; Nicolaus, T.; Gaub, H. E.; Nash, M. A. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. Biomacromolecules 2016, 17, 1330−1338.
(4) Sawano, A.; Miyawaki, A. Directed Evolution of Green Fluorescent Protein by a New Versatile PCR Strategy for Site-Directed and Semi-Random Mutagenesis. Nucleic Acids Res. 2000, 28, E78.
(5) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS One 2008, 3, e3647.
(6) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; E, G. H.; Nash, M. A. Ultrastable Cellulosome-Adhesion Complex Tightens under Load. Nat. Commun. 2014, 5, 1–8.
(7) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. Protein Sci. 2013, 22, 851–858.
(8) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.; Appel, R.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protocols Handbook 2005, 571–607.
(9) Dietz, H.; Rief, M. Exploring the Energy Landscape of GFP by Single-Molecule Mechanical Experiments. Proc. Natl. Acad. Sci. U. S. A. 2004, 101, 16192–16197.
(10) Kühner, F.; Gaub, H. E. Modelling Cantilever-Based Force Spectroscopy with Polymers. Polymer 2006, 47, 2555–2563.

Chapter 7

Immobilization Strategies for Single Molecule and Bulk Assays

7.1 Summary

Advances in enzyme-based surface coupling strategies allow to modularize the immobilization procedure. Since most of the enzymes catalyzing the immobilization are orthogonal, it is possible to first fuse the protein of interest to a surface and then post-translationally ligate the binding domains. This is especially interesting for large proteins or for proteins which do not fold properly with fusion domain.
Associated publication P9 describes a modular approach for single molecule force spectroscopy (SMFS) experiments. Proteins of interest are expressed (*in vivo* and *in vitro*) with only short amino acid tags (Sortase A- and ybbR-tags) at their termini. The protein of interest is immobilized with Sfp and the ybbR-tag to a surface. At the opposite end of the protein of interest a Sortase recognition sequence is used to post-translationally fuse a binding domain. Based on this approach, it is possible to covalently immobilize a protein of interest first and then modify it with the binding handle of choice, preserving the protein fold, as it is not co-expressed with the handle. The immobilization techniques that have been adopted and optimized for SMFS experiments can be transfered to surface-based label-free bulk techniques for kinetic measurements, like SPR (Surface Plasmon Resonance) or BLI (Biolayer-Interferometry). These methods rely on immobilization techniques with reactive groups. In order to locate proteins in close proximity to the reactive groups it is necessary to either charge the protein positively *via* lowering the pH below their isoelectric point or shield the charges with the addition of high salt concentrations. Both conditions are harsh and might harm the ligand and moreover, orient proteins randomly. **Associated manuscript M1** describes more mild, enzyme-based approaches (Sortase A, OaAEP1- and Sfp) to functionalize BLI-sensors site-specifically with a protein of interest. Furthermore it is now possible to fuse proteins which are not able to be immobilized under non-specific conditions by site-specific tags.
Being able to functionalize surfaces for bulk assays with the same chemistry used for single molecule assays increases comparability.

7.2 Associated Publication P9

Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles

by

Ellis Durner, Wolfgang Ott, Michael A. Nash, and Hermann E. Gaub

published in

ACS Omega, 2 (6), 3064-3069, (2017)

ACS OMEGA

Article

http://pubs.acs.org/journal/acsodf

Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles

Ellis Durner,[†] Wolfgang Ott,[†] Michael A. Nash,[‡,§] and Hermann E. Gaub*[,†]

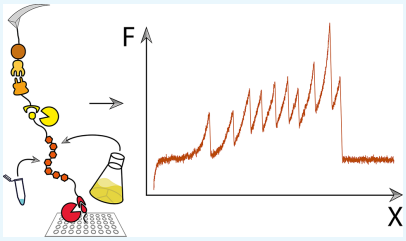
[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

[‡]Department of Chemistry, University of Basel, 4056 Basel, Switzerland

[§]Department of Biosystems Science & Engineering, Swiss Federal Institute of Technology (ETH-Zurich), 4058 Basel, Switzerland

Supporting Information

ABSTRACT: Single-molecule force spectroscopy greatly benefits from site-specific surface immobilization and specific probing with a functionalized cantilever. Here, we describe a streamlined approach to such experiments by covalently attaching mechanically stable receptors onto proteins of interest (POI) to improve pickup efficiency and specificity. This platform provides improved throughput, allows precise control over the pulling geometry, and allows for multiple constructs to be probed with the same ligand-modified cantilever. We employ two orthogonal enzymatic ligation reactions [sortase and phosphopantetheinyl transferase (Sfp)] to covalently immobilize POI to a pegylated surface and to subsequently ligate the POI to a mechanically stable dockerin domain at the protein's C-terminus for use as a high-strength pulling handle. Our configuration permits expression and folding of the POI to proceed independently from the mechanically stable receptor used for specific probing and requires only two short terminal peptide sequences (i.e., ybbR-tag and sortase C-tag). We applied this system successfully to proteins expressed using in vitro transcription and translation reactions without a protein purification step and to purified proteins expressed in *Escherichia coli*.



INTRODUCTION

In recent years, the field of single-molecule force spectroscopy (SMFS) has implemented many developments in bioconjugation to improve upon the classical approach of nonspecific pulling experiments by moving to specific, often covalent surface functionalization.[1,2] Traditionally, polyproteins are recombinantly expressed as fusion constructs framed by several repeats of marker domains of known unfolding patterns (often Ig-like domains) and nonspecifically deposited onto a surface.[3] A bare cantilever tip is then indented into the surface in an attempt to pickup and stretch single polyprotein chains on opposing ends by nonspecific adhesion. In case the number of domain unfoldings in the recorded data trace exceeds the number of domains on each side of the proteins of interest (POI), an N- to C-terminal stretching of the POI can be concluded.

In contrast to the nonspecific attachment, site-specific anchoring and probing approaches offer many advantages. They allow for homogeneous surface preparation as the immobilization geometry is defined; the usage of spacer molecules such as polyethyleneglycol (PEG) diminishes possible surface interaction effects. Drawbacks of unspecific probing—such as low-pickup efficiencies or the requirement of recombinant expression of large polyproteins—have been addressed by utilizing the receptor−ligand pairs as pulling handles to provide a specific interaction by which force can be applied to the POI. Systems such as StrepII-tag-Strep-Tactin,[4] streptavidin−biotin,[5,6] GCN4-peptide−antibody,[7] and cohesin−dockerin domains[8−10] are only a few of the interactions that have been employed for this purpose.

These pulling handles classically are genetically appended to the POI and expressed as fusion proteins. The fusion proteins are then covalently immobilized through one end of the POI and probed by ligand-functionalized cantilever-tips that recognize the respective receptor on the other end. A wide range of forces are accessible by utilizing short tags such as the StrepII-tag (116 pN at a loading rate of 4 nN s^{-1} if the tag is C-terminal and 46 pN at 4 nN s^{-1} if the tag is N-terminal[4]) and biotin (257 pN[5]), as well as that with larger handles such as the interaction between type-3 dockerin and cohesin E from *Ruminococcus flavefaciens*, reaching up to 700 pN at 100 nN s^{-1}.[10] These high-force interactions allow characterization of very stable proteins such as the unfolding of several Titin-Ig domains in series.

However, recombinant expression of a fusion between a (possibly large) POI and a large handle-protein (e.g., 29 kDa

Received: April 19, 2017
Accepted: May 26, 2017
Published: June 30, 2017

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

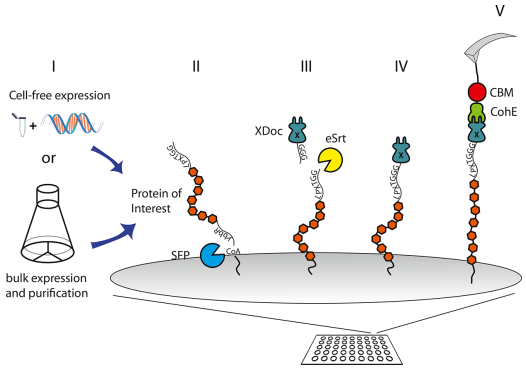


Figure 1. Schematic of the experimental setup. (I) POIs were either expressed in bulk or synthesized using a cell-free expression mix. POIs contained a ybbR-tag at the N-terminus and a sortase LPETGG tag at the C-terminus. (II) Surface-bearing PEG-coenzyme A is covalently modified with POIs via Sfp-catalyzed ligation. (III) Next, GGG-Doc is ligated to the POI at the C-terminal end using the LPETGG sortase-tag for use as a force spectroscopy pulling handle. (IV,V) Unfolding experiments are conducted by approaching and retracting a CohE-CBM-functionalized cantilever.

for CohE) can be cumbersome. The resulting fusion proteins might be insoluble or the correct folding of the POI might be affected by the presence of the fusion domains during translation and folding. Here, we utilize two orthogonal enzymatic ligation reactions to achieve sortase and phosphopantetheinyl transferase (Sfp)-mediated covalent surface attachment and post-translational modification of several POIs with dockerin handles by sortase-mediated[11,12] ligation. This allows the expression of only the protein domain of interest without risking to affect proper folding. The very robust interaction of type-3 dockerin and cohesin from *R. flavefaciens*was already shown to be functional over repeated measurements of about 24 h,[10] which is an important requirement for multiplexing atomic force microscopy (AFM) experiments.

Furthermore, we combined this technique with in vitro expression of the POI in a cell-free system. Because smaller proteins are, in general, expressed with higher yields,[13] the reduced size of the protein construct to be expressed is beneficial. This allows for a fast and easy workflow from plasmid DNA to covalently immobilized proteins containing mechanostable handles without the need for bulk expression. We anticipate that our approach will aid in highly parallel mechanical screening of mutant proteins, which benefits from the in vitro expression, obsoleting the need for protein purifications and benefitting from the enhancements in force spectroscopy throughput and robustness.

MATERIALS AND METHODS

Experimental Design. We selected Titin-Ig domains[14] and superfolder green fluorescent protein (sfGFP)[15] as the exemplary POIs for this study, as they are well-documented in the literature, and enable comparison with established methods. The POIs were cloned with a ybbR-peptide tag[16] at their N-terminus and a sortase A recognition sequence[17] LPETGG at their C-terminus. For force-spectroscopy handles, we used GGG-dockerin,[10] which was recombinantly expressed in *Escherichia coli*, purified, and ligated to the C-terminus of the POI using ligation with sortase A. While preliminary experiments were carried out with wild-type sortase A, an evolved mutant[18] was ultimately used because of its superior performance. On the cantilever side, CohE-CBM-ybbR was used and immobilized at the ybbR-tag via Sfp-catalyzed ligation.[16]

The two specific enzymatic recognition sites located on the termini of the POI ensure that only fully expressed proteins are measured in SMFS-experiments. Figure 1 shows a schematic overview of the experiment.

Cloning. Modified pET28a plasmids encoding for ybbR-His-XylanaseT6(T129C) (*Geobacillus stearothermophilus*)-Doc3 (*R. flavefaciens*), ybbR-His-sfGFP-DocI (*Clostridium thermocellum*), and Titin-Ig domains (repeats 27 to 32, repeat 34, human) were used as templates for polymerase chain reaction (PCR) with subsequent reconstitution by Gibson[19] assembly. The previously reported[18] d59 sortase(P94R/D160N/D165A/K190E/K196T) mutant was created by introducing the mutations via overlap extension PCR followed by ligating the linearized plasmid using Kinase−Ligase−DpnI (KLD) enzyme mix and KLD reaction buffer from the Q5 site-directed mutagenesis kit (New England Biolabs, MA, USA). The chemically competent *E. coli* DH5-α cells were transformed [Life Technologies GmbH, Frankfurt, Germany; 30 min on ice, 30 s heat shock at 42 °C followed by 37 °C for 1 h in a super optimal broth with catabolite repression medium] and plated on kanamycin-supplemented agar plates. For amino acid sequences, see the Supporting Information.

Protein Expression and Purification. All proteins were expressed in *NiCo21(DE3)RIPL* cells, which were cultivated in ZYM-5052 autoinduction media[20] supplemented with kanamycin and chloramphenicol. After pelleting, the cells were lysed by sonication and then centrifuged at 4 °C, 39 000 rcf for 60 min. The supernatant was filtered to 0.22 μm and applied to Ni-NTA columns (HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany). After washing with 6 column volumes of a buffer containing 25 mM Tris, pH 8.4, 300 mM NaCl, 20 mM imidazole, and 0.5 vol % Triton X-100, the bound fraction was eluted with an elution buffer containing 25 mM Tris, pH 8.4, 300 mM NaCl, and 300 mM imidazole.

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

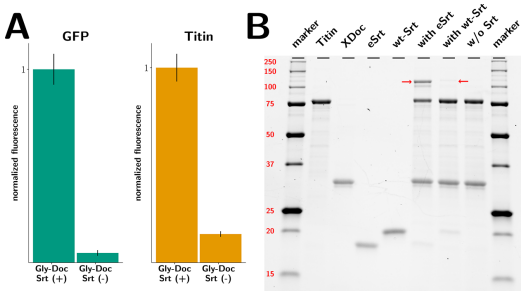


Figure 2. (A) Averaged fluorescence intensities of a CohE-CBM-ybbR-CoA647-labeled surface functionalized with ybbR-Titin-Ig-LPETGG and ybbR-sfGFP-LPETGG. Each protein was immobilized at two separate spots that were then incubated with either GGG-dockerin and sortase or GGG-dockerin but not with sortase. To test for successful ligation of dockerins, CohE-CBM-ybbR-CoA647 was allowed to bind for 10 min at 300 nM, then rinsed and imaged immediately afterward. Fluorescent intensities of each construct were normalized to the intensity of the sortase-positive spot. (B) SDS-PAGE demonstrating the ligation of GGG-dockerin to ybbR-Titin-LPETGG with wild-type sortase A (wt-Srt), pentamutant sortase A (eSrt), or no sortase as negative control. The red arrows are indicating the ligation products.

All protein solutions were concentrated using Amicon centrifugal filter units (10k MWCO, Merck KGaA, Darmstadt, Germany), followed by buffer exchange to Ca-TBS buffer (25 mM Tris, pH 7.2, 75 mM NaCl, and 1 mM $CaCl_2$) using polyacrylamide spin desalting columns. Proteins were stored at −80 °C with glycerol added to 10% (v/v). For cell-free expression, 25 μM reactions of PURExpress In Vitro Protein Synthesis Kit (New England Biolabs, Ipswich, Massachusetts) were incubated for 2 h at 37 °C, containing 300 ng plasmid DNA coding for the POIs.

In case of MGGG-His-Doc, the N-terminal methionine cleavage in *E. coli* was sufficient for the preparation of GGG-His-Doc, so that no additional protease digestion was necessary.

Surface Preparation. Surfaces and cantilevers for force spectroscopy were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, ABCR GmbH, Karlsruhe, Germany) and PEGylated with α-maleimide-hexanoic-ω-NHS PEG (NHS-PEG5000-Mal, Rapp Polymere, Tübingen, Germany) dissolved to 25 mM in 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid buffer (HEPES), 50 mM, pH 7.5 to provide suitable conditions for NHS coupling. Then, the PEGylated surfaces and cantilevers were coupled to coenzyme A (CoA, 1 mM) in sodium phosphate buffer, pH 7.2.

Silicon nitride cantilevers (BioLever mini BL-AC40TS-C2, Olympus, Tokyo, Japan) were used as force probes. Silicone masks with a grid of 1 mm-diameter holes (CultureWell Reusable Gaskets, Grace Bio-Labs, Bend, OR, USA), were applied to the CoA-functionalized glass slides to create separated incubation wells. Each purified POI was diluted to 50 μM in Ca-TBS that was supplemented with 20 mM $MgCl_2$ and Sfp enzyme was added to 10 μM. The reaction mixtures were added to the single incubation wells in the mask, enabling covalent immobilization via Sfp-catalyzed ligation of CoA and the ybbR tags.

For cell-free expression of the POIs, the cell-free expression reaction mix (PURExpress, New England Biolabs, MA, USA) was prepared to contain 100 ng of plasmid DNA. The expression mix was incubated at 37 °C for 2 h, then supplemented with Sfp enzyme to 10 μM and directly applied to the micowells without further purification. Sfp ligation reactions were performed for 2 h at room temperature. After subsequent rinsing with Ca-TBS, the wells were incubated with 100 μM GGG-Doc protein and 10 μM sortase A for 1 h. After rinsing with Ca-TBS, the silicon mask was removed, providing an array of covalently linked proteins that were modified with the dockerin handle at one end.

The sortase-catalyzed ligation reactions for Figure 2 contained 10 μM ybbR-Titin-LPETGG, 10 μM GGG-Doc, and 10 μM of either wild-type d59 sortase or the evolved pentamutant.[18] The ligation reactions were incubated for 1 h at 37 °C.

For surface functionalization tests, CohE-CBM-ybbR was labeled with CoA647 (New England Biolabs, MA, USA) in a reaction containing 25 μM CoA647, 10 μM CohE-CBM-ybbR, and 2 μM Sfp in Ca-TBS supplemented with 20 mM $MgCl_2$. The labeling reaction was incubated for 4 h at 37 °C. Free dye and Sfp enzyme were removed via preparative gel filtration with Ca-TBS as the running buffer through a Yarra 3 μm SEC-3000 (Phenomenex, Torrance, California, USA) column. Appropriate fractions [evaluated via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)] were pooled, diluted to 3 μM, supplemented with glycerol to 10% (v/v), and stored at −80 °C.

Labeled surfaces were imaged using ChemiDoc MP (Bio-Rad, Hercules, California, USA), with 625(30) nm/695(55) nm emission/excitation filters. The exposure time was 30 s; for background subtraction, a blank and clean cover slip was imaged with the same settings and an average background signal was subtracted from the measured average intensities. Intensities were quantified via Image Lab 5.2 (Bio-Rad, Hercules, California, USA) volume tool.

Single-Molecule Force Spectroscopy. All data were obtained using Ca-TBS. Measurements were taken with custom-built instruments (driven by PI-731 piezo actuators, Physik Instrumente, Germany) in conjunction with MFP-3D AFM controllers (Asylum Research, Santa Barbara, USA). Upon approaching the sample surface with the cantilever tip, the complex between cohesin/dockerin (C/D) was formed, and the cantilever was retracted from the surface at a constant

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

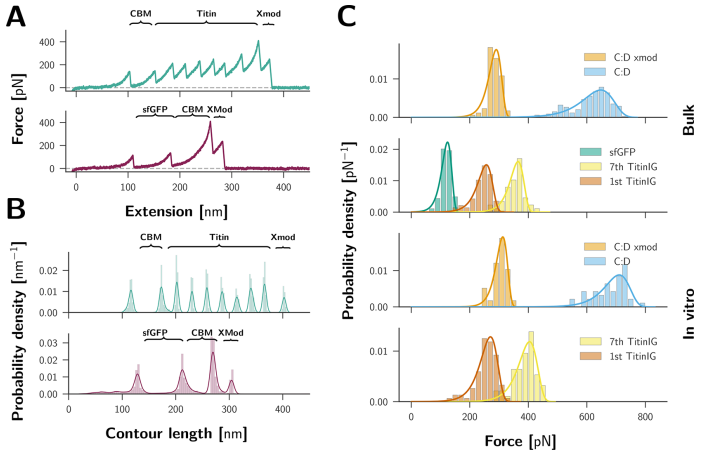


Figure 3. SMFS on Ctta-dockerin-labeled 7× Titin-Ig and sfGFP. (A) Force distance traces showing complete unfolding of the POI (Titin-Ig unfolding is shown in the upper trace, sfGFP in the lower trace). (B) Transformation of traces from (A) into contour-length space. (C) Force histograms of complex dissociation events and unfolding events of the POI: the upper two panels contain data from the bulk-expressed proteins and the lower two panels contain data from in vitro-expressed proteins. C/D complex dissociation can occur with [as in both sample traces shown in (A)] or without prior unfolding of the x-module, which is a subdomain of the dockerin, resulting in two populations of the dissociation forces. Each population was fitted with the Bell–Evans model.[23]

velocity of 800 nm s^{-1} while recoding the distance and cantilever deflection at a sampling rate of 12 500 Hz. After each force–extension curve was recorded, the sample was moved laterally by 100 nm to probe a different molecule. For data analysis, force–distance curves were transformed into contour length space using a freely rotating chain model with quantum mechanical corrections for peptide backbone stretching[21] and then sorted by contour length increments.[22] Loading rates prior to domain unfolding or complex dissociation were extracted by applying a linear fit to the last 3 nm before the respective event and then used in fitting the rupture-force histograms with the Bell–Evans model.[23]

RESULTS AND DISCUSSION

To test for successful surface functionalization, we incubated surfaces that had been prepared as described in the Materials and Methods section with fluorescently labeled cohesin. Figure 2A confirms that if sortase is ommited, no dockerin functionalization is achieved, whereas if sortase was present to perform the ligation reaction, binding of CoE-CBM-ybbR-CoA647 is observed. Figure 2B demonstrates successful ligation of GGG-dockerin to ybbR-Titin-LPETGG and illustrates the superior performance of the evolved sortase mutant d95/P94R/D160N/D165A/K190E/K196T[18] in comparison with wild-type sortase A.

Typical single-molecule force–distance unfolding patterns for the sortase-incubated spots are shown in Figure 3A. They exhibit the unbinding pattern of CohE–Doc dissociation as characterized in previous publications,[10] where dissociation can occur with or without unfolding of the dockerin subdomain called x-module. The resulting force–distance curves were transformed into contour length space and then sorted by comparing the observed unfolding increments (3B). Only curves exhibiting the 56 nm increment corresponding to a full unfolding of the CBM-domain were classified to be the result of probing a CohE–Doc complex. Furthermore, the curves were sorted to exhibit no more than one increment corresponding to the unfolding of sfGFP 79 nm and no more than seven increments corresponding to Titin-Ig unfolding 28 nM. These increments result from the added free contour length of the peptide chain upon unfolding the folded protein domains and match the previously reported values.[14,15,24] For these traces, unfolding forces of the domains of interest were histogrammed with a bin width of 20 pN (Figure 3C).

Despite its narrow tip apex, each cantilever is typically functionalized with multiple cohesin-anchors; hence, multiple receptor–ligand complexes can form if dockerin-decorated surface is densely populated. Therefore, we went for a rather sparse surface functionalization which can be tuned by the incubation times of Sfp and sortase-catalyzed ligation reactions and/or the substrate concentrations. Alternatively, cantilevers with blunter tips could be used when more interactions are desired. The achieved surface densities were in a suitable range for SMFS, sparse enough to avoid multiple interactions but dense enough to acquire good statistics. Probing attempts (1.24%) resulted in single molecule unfolding traces satisfying the outlined criteria. In total, 142 Titin-Ig and 92 sfGFP single molecule traces were obtained within 11 h of measurement with a single cantilever (spring constant: 0.093 N m^{-1}). If sortase had been omitted, no traces showing unfolding of CBM and one of the POI were recorded. For probing of in vitro-expressed Titin-Ig, 0.33% of attempts were successful, yielding 72 Titin-Ig unfoldings in 9 h of measurement, which was also probed with a single cantilever (spring constant: 0.097 N m^{-1}).

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064–3069

Figure 3C shows force histograms for unfolding events of sfGFP, the last of seven Titin-Ig domain to unfold and the complex dissociation itself. This was carried out for bulk-expressed and purified sfGFP and Titin-Ig, as well as for Titin-Ig expressed in the cell-free system. Complex dissociation events cluster into two populations that are characteristic of Doc/Coh unbinding.[10] The most probable forces at which the POI unfold are $124\left(^{+\ 24}_{-\ 16}\right)$ pN for sfGFP, $257\left(^{+\ 36}_{-\ 24}\right)$ pN for the first, and $365\left(^{+\ 33}_{-\ 22}\right)$ pN for the last Titin-Ig domain to unfold ($271\left(^{+\ 42}_{-\ 28}\right)$ and $404\left(^{+\ 45}_{-\ 30}\right)$ pN for Titin-Ig expressed in the cell-free system), the asymmetrical full widths at half maximum of the distributions are given in brackets. The most probable forces were determined by fitting each histogram of unfolding forces with the Bell−Evans model.[23]

The differences between the most probable unfolding forces observed for the POI expressed in the cell-free system and the bulk-expressed proteins are within tolerance of errors resulting from cantilever calibration.[25]

This method can be easily applied to any recombinantly expressed protein by adding the terminal peptide tags necessary for covalent surface attachment and post-translational sortase-mediated ligation. Owing to the terminal location of these tags, only nondigested and fully expressed proteins are probed. This is especially advantageous for cell-free expression systems, where the small quantity of expressed protein often makes the usually necessary affinity purification cumbersome.

CONCLUSIONS

We developed a method that enables acquisition of SMFS datasets of specifically probed and covalently immobilized single molecules. By post-translationally modifying the POI with the high-force interactions of the Coh/Doc receptor−ligand system via sortase ligation, we can probe even resilient proteins such as Titin-Ig domains with high specificity and throughput, improving on the nonspecific polyprotein method and eliminating the requirement of expressing the POI as large fusion constructs with handle domains. The modular system of post-translational attachment of the mechanostable pulling handle allowed us to probe different proteins with the same cantilever. We also applied this approach to proteins expressed in cell-free systems without further purification while still selecting for only fully expressed proteins owing to the specificity provided by the high-affinity pulling handle.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00478.

Amino acid sequences (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: gaub@lmu.de (H.E.G.).

ORCID

Hermann E. Gaub: 0000-0002-4220-6088

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the EU 7th Framework Programme NMP4-SL-2013-604530 (CellulosomePlus) and the Society in Science—the Branco Weiss Fellowship from ETH Zürich. The authors thank M. A. Jobst for the instrument control software, L. F. Milles for providing the force curve analysis software, and Thomas Nicolaus as well as Angelika Kardinal for laboratory assistance.

REFERENCES

(1) Ott, W.; Jobst, M. A.; Schoeler, C.; Gaub, H. E.; Nash, M. A. Single-molecule force spectroscopy on polyproteins and receptor−ligand complexes: The current toolbox. *J. Struct. Biol.* **2017**, *197*, 3−12.

(2) Alsteens, D.; Gaub, H. E.; Newton, R.; Pfreundschuh, M.; Gerber, C.; Müller, D. J. Atomic force microscopy-based characterization and design of biointerfaces. *Nat. Rev. Mater.* **2017**, *2*, 17008.

(3) Li, L.; Huang, H. H.-L.; Badilla, C. L.; Fernandez, J. M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **2005**, *345*, 817−826.

(4) Baumann, F.; Bauer, M. S.; Milles, L. F.; Alexandrovich, A.; Gaub, H. E.; Pippig, D. A. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.* **2016**, *11*, 89−94.

(5) Moy, V. T.; Florin, E.-L.; Gaub, H. E. Intermolecular forces and energies between ligands and receptors. *Science* **1994**, *266*, 257−259.

(6) Edwards, D. T.; Faulk, J. K.; Sanders, A. W.; Bull, M. S.; Walder, R.; LeBlanc, M.-A.; Sousa, M. C.; Perkins, T. T. Optimizing 1-μs-Resolution Single-Molecule Force Spectroscopy on a Commercial Atomic Force Microscope. *Nano Lett.* **2015**, *15*, 7091−7098.

(7) Morfill, J.; Neumann, J.; Blank, K.; Steinbach, U.; Puchner, E. M.; Gottschalk, K.-E.; Gaub, H. E. Force-Based Analysis of Multidimensional Energy Landscapes: Application of Dynamic Force Spectroscopy and Steered Molecular Dynamics Simulations to an Antibody Fragment−Peptide Complex. *J. Mol. Biol.* **2008**, *381*, 1253−1266.

(8) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20431−20436.

(9) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. From genes to protein mechanics on a chip. *Nat. Methods* **2014**, *11*, 1127−1130.

(10) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; Gaub, H. E.; Nash, M. A. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **2014**, *5*, 5635.

(11) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **1999**, *285*, 760−763.

(12) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **2013**, *8*, 1787−1799.

(13) Gagoski, D.; Polinkovsky, M. E.; Mureev, S.; Kunert, A.; Johnston, W.; Gambin, Y.; Alexandrov, K. Performance benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.* **2016**, *113*, 292−300.

(14) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997**, *276*, 1109−1112.

(15) Kufer, S. K.; Dietz, H.; Albrecht, C.; Blank, K.; Kardinal, A.; Rief, M.; Gaub, H. E. Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. *Eur. Biophys. J.* **2005**, *35*, 72−78.

(16) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15815−15820.
(17) Popp, M. W.-L.; Antos, J. M.; Ploegh, H. L. Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. In *Current Protocols in Protein Science*; John Wiley & Sons, Inc., 2009; pp 15.3.1−15.3.9.
(18) Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11399−11404.
(19) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343−345.
(20) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif.* **2005**, *41*, 207−234.
(21) Livadaru, L.; Netz, R. R.; Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules* **2003**, *36*, 3732−3744.
(22) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. Comparing proteins by their unfolding pattern. *Biophys. J.* **2008**, *95*, 426−434.
(23) Evans, E.; Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **1997**, *72*, 1541−1555.
(24) Dietz, H.; Rief, M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16192−16197.
(25) Gibson, C. T.; Smith, D. A.; Roberts, C. J. Calibration of silicon atomic force microscope cantilevers. *Nanotechnology* **2005**, *16*, 234−238.

DOI: 10.1021/acsomega.7b00478
ACS Omega 2017, 2, 3064−3069

Post-translational sortase-mediated attachment of high-strength force spectroscopy handles

Ellis Durner,[†] Wolfgang Ott,[†] Michael A. Nash,[‡,¶] and Hermann E. Gaub[*,†]

†Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

‡Department of Chemistry, University of Basel, 4056 Basel Switzerland

¶Department of Biosystems Science & Engineering, Swiss Federal Institute of Technology (ETH-Zurich), 4058 Basel Switzerland

E-mail: gaub@lmu.de

Supporting Information Available

Sequences

pET28a-MGGG-HIS-DocIII

MGGGHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTV
YNEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYY
NNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATT
VALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTF
YTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

pET28a-CohE-CBM(C63S)-HIS-ybbR

MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSG
AEGKYATTGYHIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVAS

GADDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQ
GKLMQAYFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGSVVPSTQPVTTP
PATTKPPATTIPPSDDPNAMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSS
AIDLSKLTLRYYYTVDGQKDQTFWSDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNN
ADTYLEISFTGGTLEPGAHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTA
YLNGVLVWGKEPGELKLPRSRHHHHHHGSLEVLFQGPDSLEFIASKLA

ybbR-HRV3C-HIS-sfGFP-LPETGG

MDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASSKGEELFTGVVPILVELDGDVNG
HKFSVRGEGEGDATIGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMK
RHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDG
NILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGD
GPVLLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGITHGMDELYKLPETGG

ybbR-HRV3C-HIS-Titin-LPETGG

MDSLEFIASKLALEVLFQGPLQHHHHHHPWTSASLIEVEKPLYGVEVFVGETAHFEI
ELSEPDVHGQWKLKGQPLAASPDCEIIEDGKKHILILHNCQLGMTGEVSFQAANTKS
AANLKVKELPLIFITPLSDVKVFEKDEAKFECEVSREPKTFRWLKGTQEITGDDRFE
LIKDGTKHSMVIKSAAFEDEAKYMFEAEDKHTSGKLIIEGIRLKFLTPLKDVTAKEK
ESAVFTVELSHDNIRVKWFKNDQRLHTTRSVSMQDEGKTHSITFKDLSIDDTSQIRV
EAMGMSSEAKLTVLEGDPYFTGKLQDYTGVEKDEVILQCEISKADAPVKWFKDGK
EIKPSKNAVIKADGKKRMLILKKALKSDIGQYTCDCGTDKTSGKLDIEDREIKLVRP
LHSVEVMETETARFETEISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSLVLHNCR
LDQTGGVDFQAANVKSSAHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYEDIP
VEWYLKGKKLEPSDKVVPRSEGKVHTLTLRDVKLEDAGEVQLTAKDFKTHANLFV
KAPHVEFLRPLTDLQVREKEMARFECELSRENAKVKWFKDGAEIKKGKKYDIISKG
AVRILVINKCLLDDEAEYSCEVRTARTSGMLTVLECLPETGG

7.3 Associated Manuscript M1

Site-Specific, Enzyme-Mediated Surface-Immobilization for Label-Free Binding Assays

by

Wolfgang Ott*, Ellis Durner*, Hermann E. Gaub

*(*contributed equally)*

in preparation

Enzyme-Mediated, Site-Specific Surface Immobilization for Label-Free Binding Assays

Wolfgang Ott[a,b,⊥], Ellis Durner[a,⊥], Hermann E. Gaub[a,#]

[a] Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität, 80799 Munich, Germany.

[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, 81377 Munich, Germany.

[⊥] These authors contributed equally to this work

[#] Corresponding author: gaub@lmu.de

KEYWORDS: Protein Conjugation, Sortase A, Sfp, OaAEP1, Biolayer-Interferometry, Site-Specific Protein Immobilization, Label-Free

Abstract

Interaction of proteins within a complex network appear chaotic at first glance and without direction. Bottom-up approaches enable the dissection of binding pathways and probe them individually to gain insights in their dynamic behaviour. Surface-bound affinity assays, like surface plasmon resonance (SPR) or biolayer interferometry (BLI) techniques, allow the characterization of individual ligand:analyte interactions. While it is critical to these methods that a permanent and uniform surface functionalization is achieved, ligand immobilization for SPR and BLI are commonly immobilized non-specifically *via* lysine residues. To provide a uniform and defined surface preparation and to achieve comparability to single molecule measurements like single molecule force spectroscopy, we here examine three different enzyme-mediated approaches to surface immobilization for bulk kinetic measurements. A 4´-phosphopantetheinyl transferase from *Bacillus subtillis*, Sortase A from *Staphylococcus aureus* and an asparaginyl endopeptidase from the plant *Oldenlandia affinis* were established for biosensor functionalization. All three enzymes were successfully employed in this study.

A major advantage of these enzymes is a homogeneous surface orientation of the molecules since only one anchor site per molecule exists. Secondly, proteins with lysine-residues in their active site can be immobilized, where amine-based conjugation could potentially interfere with their functionality. Thirdly, the reaction conditions are mild and do not require harsh pH or salt conditions, and work also for proteins which are not able to be immobilized *via* their amine-groups. And lastly, due to the versatility of the utilized tags it is possible to design and produce protein constructs in a single batch, which can then be probed in complementary assays.

Introduction

Complex formation between biomolecules are fundamental to the function and organization of any organism. Their association and dissociation kinetics govern the spatial arrangement of its constituents as well as they induce or inhibit signaling pathways. To study the binding properties of each of these isolated complexes *in vitro,* numerous assays were developed during the last decades. Many methods like Isothermal Titration Calorimetry (ITC)[1] determine only the equilibrium constant (besides parameter ΔG, ΔH, ΔS) but fail to deliver any information about the kinetics of the binding. So far, only surface-based biophysical techniques are well established in the scientific communities to probe on- and off rates of ligand and analyte systems. The two most commonly used methods are based on surface plasmon resonance (SPR)[2] or biolayer interferometry (BLI)[3,4]. To probe any ligand-analyte system reliably, it is necessary to stably anchor one of the binding partners onto a surface. In general, the smaller ligand is preferably linked to the surface, as this enhances resolution and sensitivity due to the thinner surface layer prior to association and larger signal upon binding.

Several methods for biosensor functionalization have been established, each with different strengths and disadvantages[5].

The most commonly used method for surface immobilization utilizes covalent linking *via* amine-[6], carboxyl-[7], or thiol-groups[8]. This way, accessible side chains of corresponding amino acids can be employed to covalently link the ligand to a surface. Covalent linkage of ligands allows for harsh regeneration conditions as well as it enables measuring complexes with extremely low off-rates, since the ligands themselves cannot dissociate. However, proteins mostly contain more than one reactive residue, which leads to inhomogeneous surface anchoring. Consequently, sensorgrams represent the response of a superposition of multiple populations of differently attached ligands. Different attachment sites may strongly influence binding kinetics solely because of molecule orientation, but especially when ligands have been immobilized *via* residues close to its binding sites, binding can be affected or even prevented (**Figure 1 A**)[9].

In this study, we expand the toolbox for biosensor functionalization by transferring and adapting advances in enzyme-based protein modification strategies.

We investigate three different enzyme-mediated pull-down methods to site-specifically and covalently link ligands *via* small recognition peptide tags to the sensor surface, hence achieving homogeneous loading of a surface (**Figure 1 B**). This enables us to link the ligand of interest to a biosensor interface in very mild reaction conditions while using only low micromolar quantities of ligand. This stands in great contrast to the non-specific attachment, which depends on a preconcentration step, where pH buffer conditions must be chosen such that sensor surface and ligands are predominantly oppositely charged, thereby attracting each other, which leads to a strong local concentration increase at the sensor surface. The close proximity of ligand to the surface is necessary to facilitate the reaction between reactive groups of sensor and ligand, like EDC/NHS-activated carboxyl-groups and amine-groups.

Low ionic strength of the utilized buffer is a prerequisite for surface preconcentration, since salt concentrations as they are commonly used in physiological buffers screen surface charges to an extent which prevents preconcentration. As a result, the chosen buffer conditions for

preconcentration might cause unfolding issues, which could lead to aggregation[10]. Also, proteins with a low isoelectric point might not be able to be fully protonated and remain negatively charged.

We chose biolayer interferometry (BLI) as method development platform because of its fast and flexible assay format. The BLI's sensors dip subsequently into wells of a 96 plate, allowing a high throughput of reagent sampling and reuse of stable reagents. Also, sample handling is somewhat easier, compared to an SPR's microfluidic system, because no liquids need to be pumped. Hence, BLI is a good choice for method development.
The underlying principle of a BLI makes use of light reflection at interfaces between media of different optical densities, analyzing the spectral shift of interference signals upon binding (and thus modifying the optical path length) to the sensor[11,12]. From this, a signal is reconstructed which changes whenever binding/unbinding to the sensor fiber occurs (**Figure 1, C**). While the presented results were obtained with a biolayer interferometer, the methods should be readily transferable to SPR assays, as the surface immobilization relies on the same chemistry.

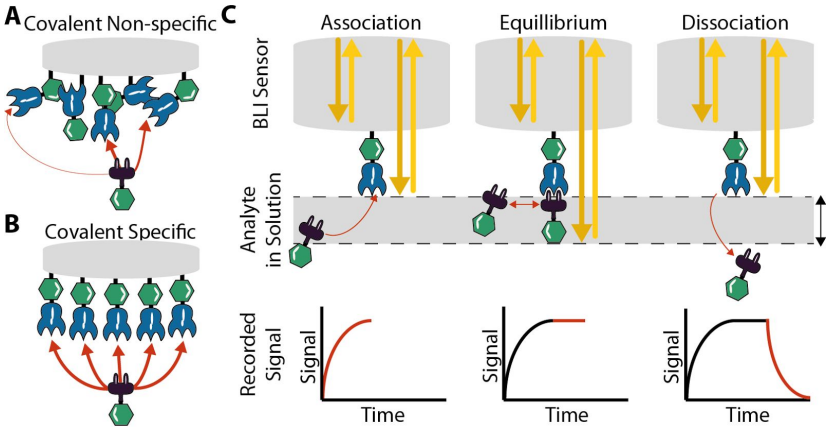


Figure 1. Schematic of BLI Kinetics. (A) illustrates non-specific immobilization of the ligand on the sensor in different geometries because of several accessible amine-groups. **(B)** sketches the specific and site-directed immobilization of a ligand to a biosensor. All ligands are homogeneously orientated. The red arrows hint to possible different binding probabilities (non-specific immobilization), whereas for the specific attachment every binding occurs about the same probability. **(C)** shows the principle of a BLI kinetic experiment. A ligand-functionalized biosensor immerse into an analyte solution. The increasing signal shows binding of the system. After the sensor is saturated as much analytes associate as dissociate - the system is in an equilibrium, thus the signal does not change. The sensor is moved to a buffer solution, the ligand starts to dissociate and the detected signal decreases again.

The employed enzymes in this study are a 4´-phosphopantetheinyl transferase (Sfp), an evolved Sortase A (d59SrtA, P94R/D160N/D165A/K190E/K196T)[13], and an engineered asparaginyl endopeptidase from the plant *Oldenlandia affinis* OaAEP1(C247A). All of these

enzymes recognize a specific amino acid sequence (tag) and covalently attach the tags to other amino acid sequences (Srt A[14] and OaAEP1[15,16]) or to Coenzyme A (CoA) (Sfp)[17]. In case of Sortase A and OaAEP1 the tags have to be at the very termini of the protein, whereas the ybbR-tag (11 amino acids) for Sfp can be also internally (if accessible) since its ligation mechanism does not rely on peptidase activity. These tags can be fused to proteins and employed in surface pull-down strategies.

Advances in bioconjugation methods and optimization of ligating enzymes[18] made their use feasible in many applications. The presented enzymes are used for post-translational protein modifications, especially for protein labeling strategies[19]. Low micromolar amounts of protein and, *i.e.* a dye, allow these reactions to proceed. Strategies moved from the classical amine-, thiol- or carboxyl-group labeling to the use of site-specific tags[20]. This way, the stoichiometry of the labeling of proteins is precisely defined, and therefore, quantification of dynamics down to the single molecule level are made possible.

Sfp- and Sortase A-based chemistry is already well established in single molecule force spectroscopy[21–23]. Also, for high affinity complexes it is necessary to have a reliable and covalent surface chemistry. The high rupture force of up to 700 pN (at 100 nN s^{-1}) of the cohesin dockerin type III complex[24] requires covalent surface pull-down strategies to prevent detachment from the surface. The enzyme-mediated approach provides covalent and site-specific tethering, yielding in a homogeneous orientation of molecules on a surface at the same time.

Since the employed tags are very small in size, they should not influence the overall functionality of the protein. This makes it very attractive to use same constructs characterized in a surface based assay as well as for other bulk and single molecule studies. For a ligand composed of a binding domain and a catalytic or characteristic domain, recombinant expression ensures that only fully translated proteins are probed in the assay. Since the immobilization tag is ideally located at one terminus and the binding domain at the opposite terminus of the protein [23].

Especially for a combined approach of single molecule studies and bulk characterization of protein complexes, these small tags are very advantageous. They can be used for post-translational protein modifications, *i.e.* attachment of a fluorescent dye[25], a protein binding domain[22,26] or as a pull-down technique[23]. This way, label-free and label-dependent techniques can be used with the same batch of proteins.

Work on non-covalent but site-specific antibody orientation in SPR measurements, *via* Protein A, pull-down, has been done before[27,28]. Depending on the antibody, a difference between specific pull-down strategies and non-specific adsorption could be detected. Hence, investigation and improvement of direct surface coupling strategies should be pursued.

Results and Discussion

In order to establish enzyme-mediated pull-down strategies for surface-based assays the well-known cohesin dockerin type III interaction of *Ruminococcus flavefaciens* was chosen as model interaction. This cohesin dockerin pair was already characterized in bulk studies[29] as well as single molecule studies[22,24,26,30,31].

Non-Specific Sensor Immobilization. A cohesin construct was diluted in different buffers with varying pH, below the cohesin's isoelectric point. None of the conditions could successfully immobilize functional cohesin domains. Cohesin also contains lysines in its binding pocket, hence it is possible that the immobilization signal produced was based on these (**Figure 2, Blue, Red and Green Trace**). A second approach with high salt conditions to shield electrostatic interactions also was not viable to bind enough cohesin to the surface (**Figure 2, Purple Trace**).

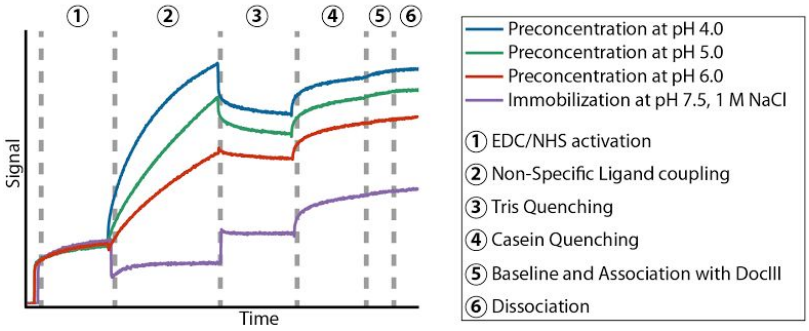


Figure 2. Sensorgram for the non-specific immobilization. In order to react amine-groups to the BLI-sensor it needs to be activated with EDC/NHS **(1).** Different non-specific immobilization techniques were probed **(2)**: 0.2 nM cohesin was diluted in 10 mM Na-Acetate buffers with a pH range of 4 - 6 (below the pI of the cohesin) to enable electrostatic attraction. In purple an alternative approach is illustrated: shielding all electrostatic interactions by adding 1 M of NaCl and increasing the cohesin concentration to 5 µM. **(3)** Tris-Quecking followed to disable all remaining, reactive EDC/NHS-groups. **(4)** Casein passivation followed to inhibit non-specific interaction of sensor with analyte. **(5 and 6)** shows negligible association and dissociation of dockerin.

Site-Specific Sensor Immobilization. We found that all three enzymes are feasible to catalyze surface attachment of ligands. Basis for all of the specific pull-down experiments were amine-reactive sensors, which contain carboxyl-groups. Carboxyl-groups can be activated with EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) and NHS (N-Hydroxysuccinimide) to create amine-reactive NHS-esters.

Sfp chemistry. In order to immobilize ybbR-tagged proteins with the magnesium-dependent enzyme Sfp, sensors need to be functionalized with its substrate Coenzyme A (CoA). Sfp catalyzes a covalent, irreversible reaction between CoA and the serine residue of a ybbR-tag (DSLEFIASKLA). Since only amine reactive sensors are available one has to make use of the amine-thiol crosslinking reagent PDEA (2-(2-pyridinyldithio)ethaneamine hydrochloride). PDEA contains an amine-group and a disulfide-bond with a leaving group facilitating thiol exchange

with other thiols. PDEA binds with its amine-group to the EDC/NHS-activated surface and in a subsequent step is able to bind CoA *via* thiol exchange, since CoA contains a thiol-group not necessary for the Sfp-reaction (**Figure 3, left**). The CoA-Biosensor is now dipped into the well, containing Sfp and ybbR-tagged protein to catalyze the sensor loading reaction.

Sortase A chemistry. The calcium-dependent Srt A needs peptide-functionalized biosensors (n-terminal ligand coupling: -LPETGG, c-terminal ligand coupling: GGG-). Hence, a specific peptide is linked to a EDC/NHS activated surface either *via* its n-terminal amine (LPETGG, **Figure 3, middle**) or *via* a cysteine and the above described PDEA-approach (GGG).
LPETGG does not contain any amines besides its n-terminal one, in order to improve reactivity and to increase the isoelectric point of the peptide (easier preconcentration) the linker sequence KKGSGSGS was added to the peptide (KKGSGSGS-LPETGG). After functionalization of the BLI-sensor with the peptide SrtA is able to link a polyglycine (n-Terminus) with a glycine from the c-terminal tag (LPETGG). It is important to note, that this covalent connection is able to be recleaved again by SrtA. Since the reaction intermediate LPET・is prone to hydrolysis, functionalization can decay again for prolonged reaction times when sensors presenting the c-terminal recognition sequence LPETGG are used.

OaAEP1 chemistry. The recent discovered and characterized OaAEP1 recognizes the N-tag (GLP) and covalently links it to a (NGL) at the c-terminus [15,16]. Similarly to the Sortase A chemistry a KKGSGSGS-NGL-peptide is fused to a EDC/NHS-activated surface (**Figure 3, right**). OaAEP1 can now catalyze the GLP-cohesin attachment to the sensor. While OaAEP1 is a transpeptidase as well, production of shunt product due to hydrolysis seems to occur at a much slower rate when compared to Sortase A, as we could not observe any degradation of the functionalization even at long reaction times.

All three sensors are then passivated with Casein, it should be noted that other protein systems might need different passivation agents. The passivated, specific reactive sensors can then be incubated with the ligand protein of interest with the appropriate tag and the catalytic enzyme of choice to attach the ligands covalently and homogeneously to the sensor.
Sortase A and OaAEP1 based chemistries can also be inverted with peptides containing a cysteine at the c-terminus and relying on the PDEA approach (See Supplemental Information, Figure S3).

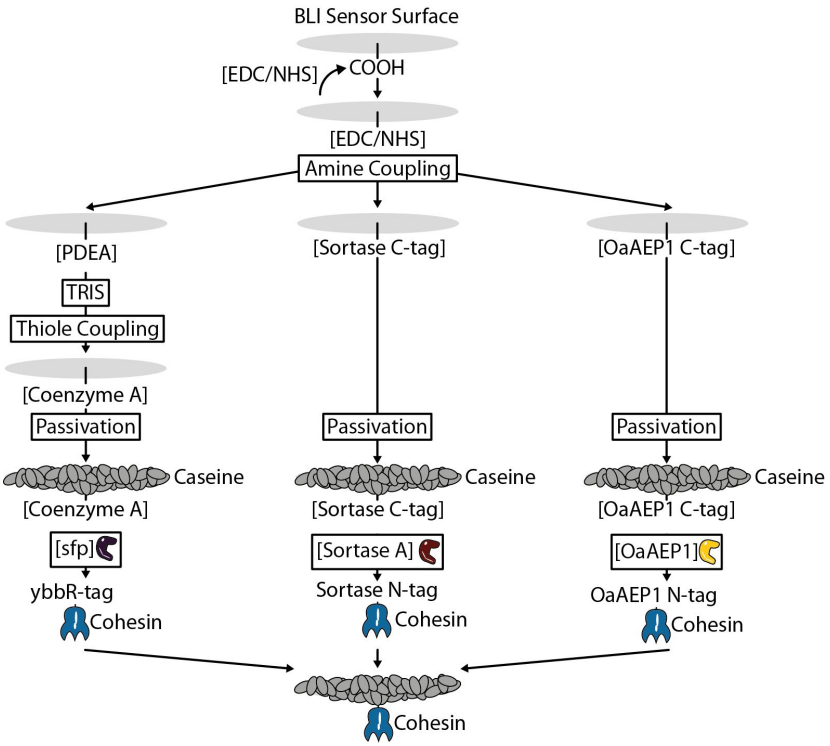


Figure 3. Overview of the different site-specific immobilization techniques. Three different enzyme-catalyzed techniques, employed in this study are shown. The sensors were passivated with Casein after the specific functionalization steps. **Left:** Sfp catalyzes the reaction between ybbR-tag of a cohesin and Coenzyme A. PDEA first reacts with the EDC/NHS activated sensor *via* its amine-group. It also contains a reactive thiol-group, which can be used for Coenzyme A functionalization. **Middle:** OaAEP1 recognizes the C-terminal amino acids NGL and fuses it to cohesins containing the N-terminal amino acids GLP. EDC/NHS activated sensors were reacted with the amine-groups of a KK-GSGS-NGL peptide. **Right:** Sortase A links C-terminal LPETGG with N-terminal GGG. In the shown case a KK-GSGSGS-LPETGG peptide was reacted with the EDC/NHS sensor to any of its three N-terminal primary amines.

Surface-Functionalization in Real time Detectable. Figure 4 shows an example signal trace for the Sfp functionalization. A sensor, in MES-buffer equilibrated, dips into the EDC/NHS solution (**Figure 4, 1**). An increase in signal can be detected. The following PDEA step also shows binding (**Figure 4, 2**). Next an amine quenching is necessary, because CoA also contains amine-groups. When linked *via* its amine-groups, CoA cannot be ligated to the ybbR-tag (**Figure 4, 3**). The thiol exchange reaction between CoA and PDEA can now take place. It is noteworthy that CoA does not generate any signal upon binding at concentrations below 20 mM (**Figure 4, 4**). Next, the sensor needs to be passivated with a passivation agent

suited for the ligand-analyte pair. In our case, Casein worked best (**Figure 4, 5**). The site-specific Sfp reaction can now take place. In green a trace is shown with added Sfp, in red only the ligand was added, Sfp was omitted (**Figure 4, 6**). The sensor of the red trace is dipped into a well with only cohesin, hence the small increase in signal is due to non-specific binding. We therefore can correct for the content of non-specific binding during the enzyme-catalyzed reaction (**Figure 3, 6, Green Trace**). During the association and dissociation of the dockerin to and from cohesin (**Figure 3, 7 and 8**) a very small signal of interaction between non-specifically adsorbed cohesin (red trace) is detected. The positive binding signal can be later corrected with this value. The almost vertical jumps in the signal traces are due to changing buffer conditions which cause an abrupt change in the signal.

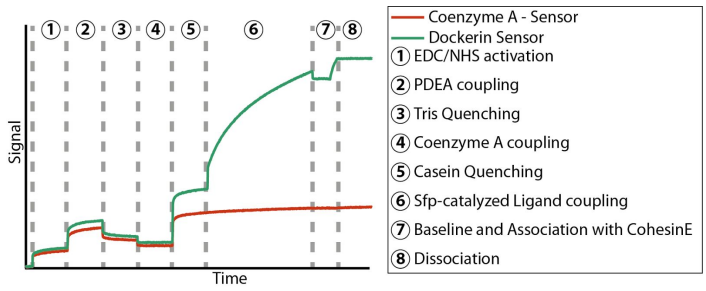


Figure 4. Exemplary Sensorgram for the Sfp-based chemistry. First the carboxyl-group is activated with EDC/NHS **(1)** which reacts with the amine-group of PDEA **(2)**. **(3)** Tris quenches all the unreacted, but active EDC/NHS-groups. Coenzyme A **(4)** shows no increase in the sensorgram but nevertheless links to the PDEA. **(6)** shows the specific Sfp reaction (green) and the non-specific adsorption of the ybbR-dockerin to the sensor (red). After this functionalization the sensor is ready to measure association and dissociation of analytes **(7 and 8)**.

In order to exclude any binding caused by potentially non-specifically adsorbed Sfp and the ybbR-HIS-Xmod-DocIII further negative controls were made (**See Supplemental Information, Figure S1 and S2**). A CoA-modified sensor was dipped into a Sfp containing mix. After dissociation and equilibration of the sensor in measuring buffer the sensor was moved into a cohesin containing mix. No binding was observed at the employed concentrations (5 µM Sfp and 1 µM cohesin), which are above the determined K_D of 20 nM[29]. Hence, the generated interaction of the negative sensor (**Figure 4, Red Trace**) is based on non-specific cohesin adsorption and the binding of dockerin.

For the CoA-chemistry, as well as when peptides presenting the n-terminal tags for SrtA or OaAEP1 ligation are used, sensor surfaces are modified *via* disulfide exchange to the PDEA crosslinker molecule. This makes it possible to regenerate the sensor by adding a reducing agent. However, it is not possible to measure the binding kinetics under reducing conditions, otherwise the ligand would detach from the sensor. If reducing conditions are required, then

sensors can be functionalized *via* the c-terminal tags for Srt A or OaAEP1 *via* the amine-groups of the peptides.

The sensors of the BLI can also be functionalized offline, once the reaction conditions are optimized with the optical readout. Offline functionalization means, that the sensors are manually moved from one well to the next one in the sensor tray. This is especially advantageous for the Octet K2, since it only can operate two sensors at a time.

Specific immobilization of cohesin possible. Cohesin could be attached site-specifically and biologically active with Sortase A and Sfp, in contrast to the non-specific immobilization. **Figure 5** shows the results of the experiment, with all steps necessary to load a ligand and probe its analyte binding. In case of SrtA either a dockerin (**Figure 5, A**) or a cohesin (**Figure 5, B**) was immobilized to probe both immobilization strategies (n- or c-terminal anchoring). Cohesin and dockerin were chosen, because their native tethering orientation is mirrored (n-terminal: dockerin, c-terminal: cohesin). Depending on the immobilized molecule a different unbinding behaviour can be observed. This might be caused due to different binding and unbinding geometries or a varying interaction of sensor and ligand. The n-terminally immobilized dockerin with OaAEP1 (**Figure 5 C**), deviates slightly in its unbinding behavior from the Sortase A tethered dockerin.
Figure 5 D depicts the loading of cohesin to BLI-sensors with Sfp-based chemistry. In this case the c-terminal cohesin dissociation varies from the c-terminal SrtA-based dissociation, which might be caused by a higher degree of flexibility, due to the PDEA-based surface chemistry.

In general it is now possible to anchor ligands to surfaces without relying on harsh conditions and cumbersome screening runs to find and optimize ligand loading conditions, however further optimization is necessary since all strategies should return same (un)binding behaviour. Once optimized mild anchoring strategies are available and investigating differences in binding geometries is possible.

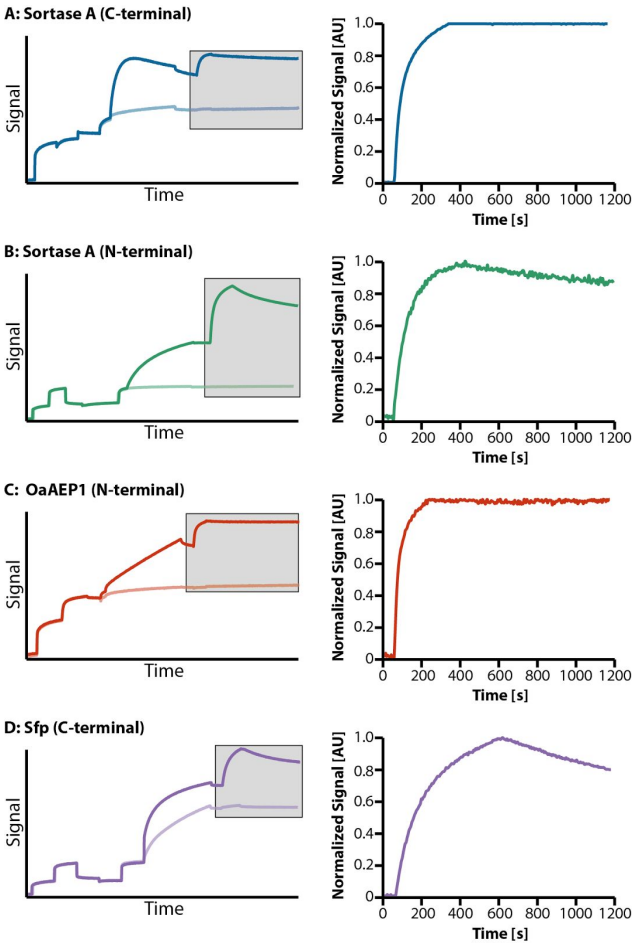


Figure 5. Binding Kinetics of the different surface chemistries. A complete sensorgram of each surface chemistry is shown. Dark colors represent the positive signal (enzyme + ligand in loading step), light colors the negative control (only ligand in loading step). The grey box highlights association and dissociation phase.The right graph shows corrected and normalized sensor signal over time of association and dissociation (positive signal subtracted with negative signal and binding signal normalized to 1). **(A)** shows the Sortase-A-based GGG-dockerin functionalization. **(B)** represents the sensorgram of the inverse Sortase A coupling with cohesin-LPETGG **(C)** Sensorgram of OaAEP1-based chemistry with GLP-dockerin. **(D)** shows in purple the Sfp-based chemistry with cohesin-ybbR.

Outlook

Single molecule studies rely on site-specific surface anchors. With the here presented strategies one and the same constructs can be used in bulk and single molecule assays, an important step towards comparability. Independent of the functionalization strategy, all ligands could be loaded to the sensor, in contrast to all non-specific approaches. Ensemble averaged results obtained from single molecule experiments can be used to interpret bulk assays. Also, it is easy to observe and optimize different coupling steps with the BLI and transfer the so obtained insights to the single molecule sample preparation.

The here presented techniques are shown for sensor modification of BLI. They also can be easily adopted for SPR chips, since both rely on the same chemistries. While passivation, reaction times and concentrations of the compounds may likely require optimization for each ligand, the enzymatic immobilization approaches greatly facilitate surface immobilization as no buffer conditions suitable for preconcentration need to be found. This also counts for new receptor ligand interactions, where buffer conditions may need optimization. Overall, the site-directed immobilization techniques present a viable, easily implementable alternative to the classical non-specific approach.

Methods

All materials employed in this study were at least from analytical purity grade and purchased from Sigma Aldrich Chemie GmbH (Taufkirchen, Germany) and Carl Roth + Co. KG (Karlsruhe, Germany).
All buffers were filtered (0.22 µm) and degassed *via* sonication. The pH of the buffers were adjusted at room temperature.

Cloning.
All constructs were subcloned into a modified pET28a vector with Gibson Assembly[32]. In order to perform Gibson Assembly, the inserts were ordered as a gene string with overlapping sequences to the plasmid, up- and downstream of the insert or amplified with primers containing the overlaps. The reaction was performed at 50°C for 1 hr (10 µl 2x HiFi MasterMix; New England Biolabs, Ipswich, MA, USA, mixed with 0.1 nmol of vector and 0.2 nmol of insert). Later modifications were done with a plasmid linearization reaction *via* PCR and recircularization reaction (4.5 ml of PCR product, 1 µl ATP (10 mM, Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.5 µl PEG-6000 (Thermo Fisher Scientific Inc.), 1 µl CutSmart buffer (10x, New England Biolabs), 1 µl T4 Polynucleotidekinase (Thermo Fisher Scientific Inc.), 1 µl *Dpn*I (Thermo Fisher Scientific Inc.), 1 µl T4 Ligase (10 U/µl,Thermo Fisher Scientific Inc.). This reaction was performed 15 Mins at 37°C and 45 Mins at 22°C.
DH5α cells (Life Technologies GmbH, Frankfurt, Germany) were thawed on ice, 1 µl of the reaction mix was added to the cells to transform them. After 30 Mins on ice a heat shock at 42°C for 1 Min was done. Finally the cells grew shaking (850 rpm) for 1 hr at 37°C in 1 ml of SOC Medium. Usually 100 µl of the transformed culture was used to streak on a LB-Kanamycin containing plate. It was incubated overnight at 37°C. Plasmids of a small amount of clones (less than 5) were amplified and sent to sequencing to verify their sequence.

The gene for OaAEP1 was ordered *via* gene string (Thermo Fisher Scientific Inc.) and subcloned *via* Gibson Assembly. Sortase A was cloned and modified as described by Durner *et al*[22].

CohE-HIS-ybbR, ybbR-HIS-Xmod-DocIII, CohE-HIS-LPETGG, MGGG-HIS-Xmod-DocIII and MGLP-HIS-Xmod-DocIII were created with the above described linearization and recircularization reaction. The underlying constructs were ybbR-HIS-CBM-Xmod-DocIII and CohE-CBM-HIS-ybbR published by Schoeler *et al*[24].

Protein Expression and Purification.
NiCo21(DE3) (New England Biolabs) cells were transformed with the appropriate plasmid (50 ng), and incubated overnight at 37°C on LB-Agar plates with kanamycin (50 ng/µl). One clone was inoculated in 5 ml LB-Kanamycin liquid culture (37°C, 200 rpm, 12-16 hrs). This starter culture was used to inoculate a larger 400 ml autoinduction media culture (100 ng/µl kanamycin)[33]. Usually the cultures were incubated for 24 hrs (4 hrs 37°C, 20 hrs 18°C; 120

rpm). Cells were harvested *via* centrifugation (6500 g, 20 Mins, 4°C), the supernatant discarded and the pellets frozen at -80°C until further use.
The frozen pellets were resolubilized in 40 ml lysis buffer (50 mM Tris (tris(hydroxymethyl)aminomethane)-HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 10 % (v/v) glycerol, 0.1 % (v/v) Triton X-100 and supplemented with 10 µg/ml DNase I and 100 µg/ml lysozyme). To enhance the chemical lysis the cells were sonicated with a sonication lance twice for 7 Mins on ice (50 % Power, 50 % Cycle; Bandelin Sonoplus GM 70, tip: Bandelin Sonoplus MS 73, Berlin, Germany). The cell lysate was centrifuged (30.000 g, 4°C) for 1 hr. The supernatant was filtrated (0.45 µm) and applied to a Ni-NTA column (5 ml HisTrap FF, GE Healthcare Europe GmbH, Freiburg, Germany), since all proteins contained a HIS-tag. The protein was eluted after washing (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 20 mM imidazole, 0.25 % (v/v) Tween 20, 10 % (v/v) glycerol) with 6 column volumes with elution buffer (25 mM Tris-HCl pH 7.8, 500 mM NaCl, 300 mM imidazole, 0.25 % (v/v) Tween 20 (v/v), 10% (v/v) glycerol). The eluted proteins were checked on a SDS-gel for digestion and purity.
The buffer of the protein solution was exchanged with ZebaSpin columns (Zeba spin desalting columns 7K, Thermo Fisher Scientific Inc.) to 50 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$ and 20 % (v/v) of glycerol was added. Small aliquots were flash frozen in liquid nitrogen and stored at -80°C.

Production of OaAEP1.
Expression and Purification of OaAEP1 roughly followed recently published protocols[15,16].
SHuffle® T7 Competent *E. coli* (New England Biolabs) were transformed with a plasmid encoding for HIS-Ub-OaAEP1(C247A) and plated on LB-Agar plates with kanamycin (50 ng/µl).
A 5 ml LB-Kanamycin liquid culture was inoculated with one colony (30°C, 200 rpm, 12-16 hrs).
A 1 l culture of ZY-505 medium was inoculated with this preculture and grown to an OD of 4, when expression was induced by adding IPTG to a concentration of 100 µM and the temperature was lowered to 16°C. After 18 hours, cells were harvested by centrifugation (6500 g, 10 minutes, 4°C), the supernatant discarded and the pellets frozen at -80°C until purification.
For purification, the frozen pellet was resolubilized, lysed and applied to a Ni-NTA column as described above. Eluate containing HIS-Ub-OaAEP1(C247A) was pooled and its pH was adjusted to 4 by adding acetic acid, enabling the autocatalytic activation during which both termini of OaAEP1 are cleaved. After 10 hours at 37°C, precipitate was removed by centrifugation and the supernatant was concentrated using Amicon centrifugal filter units (10k MWCO, Merck KGaA, Darmstadt, Germany).
To remove cleaved fragments and to exchange the buffer (to 100 mM NaAcetate, 50 mM NaCl, pH 5.5), the concentrate was applied to a Superdex 75 Increase column (GE Healthcare Europe GmbH). Fractions containing cleaved OaAEP1(C247A) were pooled and supplemented with glycerol to 10% (v/v), then flash-frozen in liquid nitrogen and stored at -80°C.

BLI Sensor Modification.

Basic sensor preparation
For all surface immobilisation strategies the AR2G - Amine Reactive 2nd Generation sensors (Pall ForteBio LLC, Fremont, CA, USA) were used. Sensors were equilibrated in 50 mM MES (2-(*N*-morpholino)ethanesulfonic acid) buffer, pH 6.0 for at least 15 minutes at room temperature.
After proper rehydratation, they were activated to an amine reactive state by incubating the sensors with a mixture of 50 mM EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) and 50 mM NHS (N-Hydroxysuccinimide) (in 50 mM MES buffer, pH 6.0) for 10 minutes.
Sensors were moved into 50 mM HEPES ((4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)), pH 7.5 after, specific or non-specific, loading of the sensor with the ligand. This ensured same conditions for the recording of the binding kinetics.

Non-specific Sensor Immobilization
The EDC/NHS-activated sensor was immobilized with GGG-cohesin-HIS (0.2 µM) in 10 mM Na-Acetate buffer, pH 4.0, 5.0, 6.0 for 10 minutes.
Also a non preconcentration run was tested with 5 µM GGG-cohesin-HIS in 100 mM HEPES, pH 7.5 and 1 M NaCl. Two subsequent quenching steps followed with 100 mM Tris-HCl (pH 8.5) and measuring buffer (25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Casein, 0.1 % (v/v) Tween-20) for 10 minutes each.

Specific Sensor Immobilization

Basic sensor preparation
If the specific protocol relied on a disulfide exchange reaction, the now amine reactive sensors were dipped for 10 minutes into a 40 mM solution of PDEA (2-(2-pyridinyldithio) ethaneamine hydrochloride, GE Healthcare Europe GmbH) (dissolved in 50 mM borate buffer, pH 8.5) to covalently immobilize the thiol reactive compound to the biosensor. In order to quench remaining reactive EDC/NHS-groups, the PDEA modified sensors were incubated for 10 minutes in 100 mM Tris-HCl, pH 8.5.

Sfp Chemistry.
Sensors were modified with PDEA as described under basic sensor preparation.
Since Coenzyme A (CoA, Merck KGaA Darmstadt, Germany) contains an accessible thiol-group, the PDEA modified sensors can react with them. 1 mM CoA in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) was fused to the sensors for 10 minutes. A final quenching step with 0.1 % (v/v) Casein (Sigma-Aldrich) passivated the remaining sensor surface against non-specific attachment.
The ligand of choice with a ybbR-tag was covalently attached to the CoA with the help of Sfp (25 µM ligand, 5 µM Sfp in 25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 0.1 % (v/v) Casein, 0.1 % (v/v) Tween-20) for 30 minutes. The sensors are now ready for kinetic binding measurements.

Sortase and OaAEP1 Chemistry.
For the peptide based Sortase or OaAEP1-mediated sensor functionalizations, two different strategies were employed. Either, peptides containing the c-terminal Sortase or OaAEP1 recognition sequences were directly coupled to the sensors *via* primary amines.
For functionalizing the sensors with the n-terminal recognition sequences, the 'adapter molecule' PDEA was employed to pull-down the peptides bearing a cysteine at their c-termini *via* a disulfide exchange to the cysteines' SH-groups.

Sortase C-tag.
Sensors were brought to an amine reactive state as described under basic sensor preparation.
Amine reactive sensors were then dipped for 10 minutes into a solution of 200 µM KKGSGSGSLPETGG peptide (GenScript, Piscataway, USA) in 10 mM Tris-HCl, pH 7.2. With two lysines located at the n-terminus, conjugation can occur to any of the three amine-groups of the peptide, the c-terminal Sortase recognition sequence is connected by a 3xGS linker. To quench any potentially remaining amine-reactivity, sensors were then incubated in 100 mM Tris-HCl, pH 8.5 for 5 minutes. To prevent non-specific adhesion in the following steps, sensors were then passivated by incubating them with 0.1 % (v/v) Casein (Sigma-Aldrich). For functionalization with the desired ligand (exhibiting the n-terminal Sortase-Tag GGG), sensors were incubated with 10 µM ligand, 1 µM Sortase enzyme and 0.1 % (v/v) Casein in 25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$) until a desired functionalization density is reached, usually for about 5 to 10 minutes.

Sortase N-tag.
Sensors were prepared to a thiol reactive state by modifying them with PDEA as described under basic sensor preparation.
Thiol reactive sensors were then loaded with GGGGGC peptide by performing a disulfide exchange reaction, replacing the PDEA's leaving group with the peptide. This reaction was performed by incubating the sensors for 20 minutes in a solution of 200 µM peptide in 10 mM sodium acetate buffer at pH 4.5.
Subsequently, sensors were then passivated by incubating them with 0.1 % (v/v) Casein (Sigma-Aldrich). For functionalization with the desired ligand (exhibiting the c-terminal Sortase-Tag LPETGG), sensors were incubated with 20 µM ligand, 1 µM Sortase enzyme and 0.1 % (v/v) Casein in 25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20) until a desired functionalization density is reached, usually for about 5 to 10 minutes.

OaAEP1 C-tag.
Sensors were prepared to an amine reactive state as described under basic sensor preparation.
Amine reactive sensors were then coupled to KKGSGSGSNGL peptides by dipping them into a solution of 200 µM peptide in 10 mM HEPES-HCl at pH 7.2.
Hereafter, potentially remaining amine reactive groups were quenched with 50 mM Tris-HCl, pH 8.5 for 5 minutes.
Sensors were then passivated with a solution of 0.1 % (v/v) Casein (Sigma-Aldrich).

For functionalization with the desired ligand (exhibiting the c-terminal OaAEP1-Tag NGL), sensors were incubated with 20 µM ligand, 1 µM OaAEP1 enzyme and 0.1 % (v/v) Casein in 25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20) until a desired functionalization density is reached, usually for about 5 to 10 minutes.

OaAEP1 N-tag.
Sensors were prepared to a thiol reactive state by modifying them with PDEA as described under basic sensor preparation.
Thiol reactive sensors were then loaded with GLPGSC peptide by performing a disulfide exchange reaction, replacing the PDEA's leaving group with the peptide. This reaction was performed by incubating the sensors for 20 minutes in a solution of 200 µM peptide in 10 mM sodium acetate buffer at pH 4.5.
Subsequently, sensors were then passivated by incubating them with 0.1 % (v/v) Casein (Sigma-Aldrich). For functionalization with the desired ligand (exhibiting the C-terminal OaAEP1-Tag NGL, sensors were incubated with 10 µM ligand, 1 µM OaAEP1 enzyme and 0.1 % (v/v) Casein in 25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Tween-20) until a desired functionalization density is reached, usually for about 5 to 10 minutes.

BLI measurements.
All measurements were done at 30°C with an Octet® K2 System (Pall ForteBio LLC.). Sensors, functionalized with the ligand, were equilibrated for 2 minutes in measurement buffer (25 mM Tris-HCl, pH 7.2, 72 mM NaCl, 1 mM $CaCl_2$, 0.1 % (v/v) Casein, 0.1 % (v/v) Tween-20). Two sensors were moved synchronously into wells with analyte. One of the two sensors was functionalized specifically with the ligand, the other was prepared as a reference sensor by ommitting the ligating enzyme during functionalization so that no ligand is presented at the sensor surface. The association was done for 2 minutes, after that the sensors moved back into the wells with measurement buffer and dissociation was recorded for more than 20 minutes.

Raw Data was exported as csv from ForteBio Data Analysis 9.0 and then fitted using a custom script written in Python (Python Software Foundation, Python Language Reference, version 2.7, available at http://www.python.org), utilizing the libraries NumPy, SciPy, and Matplotlib.
In brief, data from the reference channel was subtracted from that of the functionalized channel. Then, a single exponential decay (eq. 1) was fitted to the dissociation under optimization of k_{off} and an offset y_{off}. The so obtained off-rate was used to fit the on-rate to eq. 2 under optimization of k_{on}.

$$y = (y_0 - y_{off}) * exp(-k_{off} * (t - t_{o,dissociation})) + y_{off}$$
$$y = y_0 (1 - exp(-(conc * k_{on} + k_{off}) * (t - t_{0,association})))$$

Acknowledgements
This work was supported by the Advanced Grant "Cellufuel" of the European Research Council and the Deutsche Forschungsgemeinschaft through SFB 1032.

SUPPLEMENTAL INFORMATION

Enzyme-Mediated, Site-Specific Surface Immobilization for Label-Free Binding Assays

Wolfgang Ott[a,b,⊥], Ellis Durner[a,⊥], Hermann E. Gaub[a,#]

[a] Lehrstuhl für Angewandte Physik and Center for NanoScience, Ludwig-Maximilians-Universität, 80799 Munich, Germany.

[b] Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, 81377 Munich, Germany.

[⊥] These authors contributed equally to this work

[#] Corresponding author: gaub@lmu.de

Protein Sequences.

Enzymes

sfp-HIS

MKIYGIYMDRPLSQEENERFMTFISPEKREKCRRFYHKEDAHRTLLGDVLVRSVISRQYQLDKS
DIRFSTQEYGKPCIPDLPDAHFNISHSGRWVIGAFDSQPIGIDIEKTKPISLEIAKRFFSKTEYSDL
LAKDKDEQTDYFYHLWSMKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELPDSHSPCYIKTYE
VDPGYKMAVCAAHPDFPEDITMVSYEELLEASHHHHHH

Sortase A-HIS

MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATREQLNRGVSFAEENESLDDQNISIAGHTFID
RPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTAVGVLDEQKGKDKQLTLITCDDY
NEETGVWETRKIFVATEVKHHHHH

HIS-Ubiquitin-OaAEP1(C247A)

MHHHHHHGSGSQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRT
LSDYNIQKESTLHLVLRLRGGARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTRWAVLIAGS
KGYANYRHQAGVCHAYQILKRGGLKDENIVVFMYDDIAYNESNPRPGVIINSPHGSDVYAGVPK
DYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGAAGVIGMPSKPYLYADEL
NDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALTSTNTTESSWAYYCPAQENPPP
PEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQYHHVDKRISHASHATQYGNLKLGEEGLFVY
MGSNPANDNYTSLDGNALTPSSIVVNQRDADLLHLWEKFRKAPEGSARKEEAQTQIFKAMSHR
VHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQPLVDDWACLRSLVGTFETHCGSLSEYGMRHTRTI
ANICNAGISEEQMAEAASQACASIP

Ligand/Analyte

CohE-HIS-ybbr

MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGY
HIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTVELK
VPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYLVKA
NATYADGYIAIKAGEPHHHHHHDSLEFIASKLA

ybbR-HIS-Xmod-DocIII

MDSLEFIASKLAHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVY
NEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYYNNATLKDAEG
NDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATTVALSPSTLVGGNPESVYDD
FSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTG
DAEEE

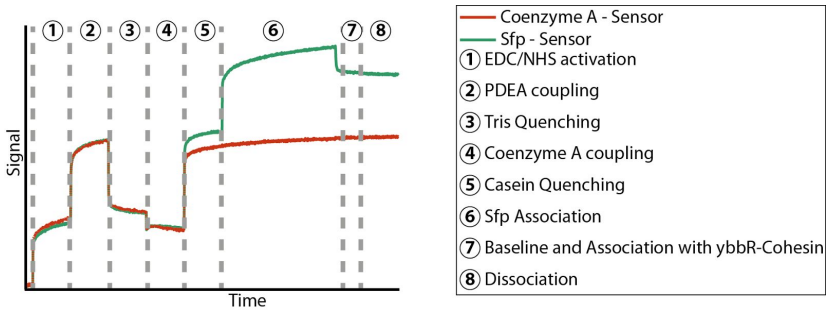
CohE-HIS-HRV3C-LPETGG
MALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGY
HIYWDERLEVVATKTGAYAKKGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTVELK
VPADAKAGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPSTDEYLVKA
NATYADGYIAIKAGEPELKLPRSRHHHHHHGSLEVLFQGPLPETGG

MGGG-HIS-Xmod-DocIII
MGGGHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVYNEGYTGD
DGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYYNNATLKDAEGNDATVTV
YIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSD
VKVDAGKELTRFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

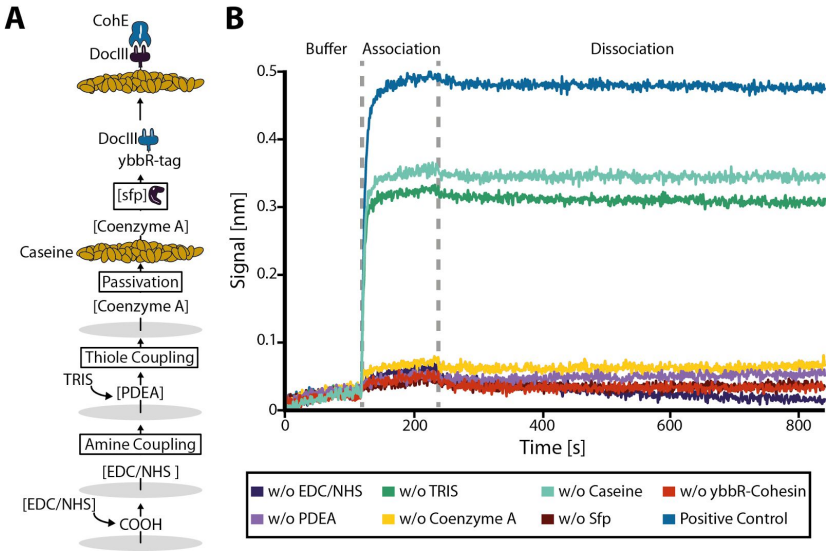
MGLP-HIS-Xmod-DocIII
MGLPHHHHHHGVVPNTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKAVLHTVYNEGYTGD
DGVAVVLREYESEPVDITAELTFGDATPANTYKAVENKFDYEIPVYYNNATLKDAEGNDATVTV
YIGLKGDTDLNNIVDGRDATATLTYYAATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSD
VKVDAGKELTRFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

Peptides

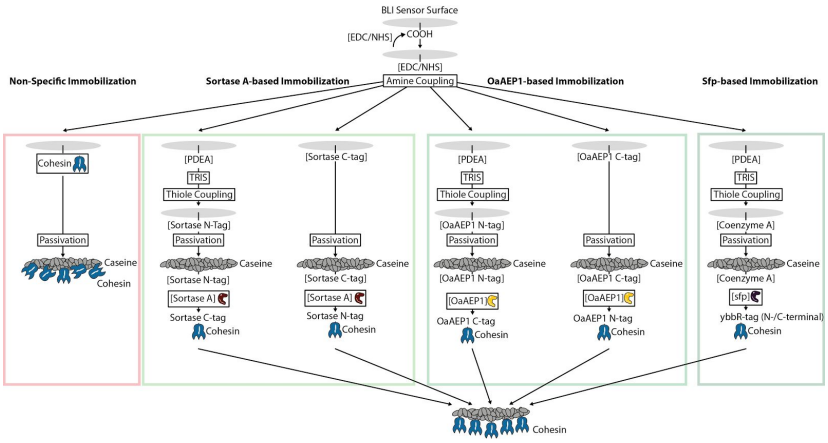
KKGSGSGSLPETGG
KKGSGSGSNGL
GLPGSC
GGGGGC



Supplemental Figure S1. Negative control for Sfp:ybbR-Interaction. First the carboxyl-group is activated with EDC/NHS (**1**) and reacts with PDEA (**2**) Tris quenching (**3**), Coenzyme A functionalization (**4**) and Casein (**5**) quenching steps followed. (**6**) shows the Sfp binding to Coenzyme A (green) and plain buffer interaction without Sfp (red). Both sensors were moved into a ybbR-cohesin solution and no binding signal was detected **(7 and 8)**.



Supplemental Figure S2. (A) schematic of the Sfp-based surface functionalization of dockerin to the BLI-sensor. First carboxyl-groups are activated with EDC/NHS and coupled with PDEA. PDEA is thiol reactive and couples in this case Coenzyme A. Coenzyme A is recognized by Sfp and covalently links a ybbR-containing dockerin. The sensor is now able to bind the corresponding cohesin. **(B)** shows association and dissociation of cohesin to the dockerin functionalized sensor. Each trace represents one control where one reagent was left out in the proper surface chemistry. Blue represents the positive control with all added substances.



Supplemental Figure S3. Overview of the all immobilization strategies. Non-specific immobilization (red box) relies on amine-groups of a ligand resulting in different orientation of the molecules on the surface.
Three different site-specific, enzyme-based strategies are shown in the green boxes:
Left:. Sortase A links C-terminal LPETGG with N-terminal GGG. Either a KK-GSGSGS-LPETGG peptide or a GGGGG-C was reacted with the EDC/NHS sensor to any of its three N-terminal primary amines of the LPETGG-peptide or *via* PDEA in case of the GGGGG-peptide.
Middle: OaAEP1 recognizes the C-terminal amino acids NGL and fuses it to proteins containing the N-terminal amino acids GLP. EDC/NHS activated sensors were reacted with the amine-groups of a KK-GSGS-NGL peptide or *via* PDEA for the GLP-C case.
Right: Sfp catalyzes the reaction between ybbR-tag of a protein and Coenzyme A. In order to enable the reaction PDEA is used to crosslink EDC/NHS-activated sensors with Coenzyme A.

References

1. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11,** 560–566 (2001).
2. Karlsson, R. SPR for molecular interaction analysis: a review of emerging application areas. *J. Mol. Recognit.* **17,** 151–161 (2004).
3. Yang, D., Singh, A., Wu, H. & Kroe-Barrett, R. Comparison of biosensor platforms in the evaluation of high affinity antibody-antigen binding kinetics. *Anal. Biochem.* **508,** 78–96 (2016).
4. Concepcion, J. *et al.* Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb. Chem. High Throughput Screen.* **12,** 791–800 (2009).
5. Oliverio, M., Perotto, S., Messina, G. C., Lovato, L. & De Angelis, F. Chemical Functionalization of Plasmonic Surface Biosensors: A Tutorial Review on Issues, Strategies, and Costs. *ACS Appl. Mater. Interfaces* **9,** 29394–29411 (2017).
6. Sheehan, J. C. & Hlaviia, J. J. The Use of Water-Soluble and Basic Carbodiimides in Peptide Synthesis. *J. Org. Chem.* **21,** 439–441 (1956).
7. Hoare, D. G. & Koshland, D. E. A Procedure for the Selective Modification of Carboxyl Groups in Proteins. *J. Am. Chem. Soc.* **88,** 2057–2058 (1966).
8. Renberg, B., Shiroyama, I., Engfeldt, T., Nygren, P.-K. & Karlström, A. E. Affibody protein capture microarrays: synthesis and evaluation of random and directed immobilization of affibody molecules. *Anal. Biochem.* **341,** 334–343 (2005).
9. Steen Redeker, E. *et al.* Protein engineering for directed immobilization. *Bioconjug. Chem.* **24,** 1761–1777 (2013).
10. Gedig, E. T. Surface Chemistry in SPR Technology. in *Handbook of Surface Plasmon*

Resonance (ed. Schasfoort, R. B. M.) 173–220 (Royal Society of Chemistry, 2017).

11. Hänel, C. & Gauglitz, G. Comparison of reflectometric interference spectroscopy with other instruments for label-free optical detection. *Anal. Bioanal. Chem.* **372,** 91–100 (2002).
12. Schmitt, H.-M., Brecht, A., Piehler, J. & Gauglitz, G. An integrated system for optical biomolecular interaction analysis. *Biosensors and Bioelectronics* **12,** 809–816 (1997).
13. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11399–11404 (2011).
14. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 13343–13348 (2014).
15. Harris, K. S. *et al.* Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6,** 10199 (2015).
16. Yang, R. *et al.* Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.6b12637
17. Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15815–15820 (2005).
18. Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7,** 876–884 (2011).
19. Milczek, E. M. Commercial Applications for Enzyme-Mediated Protein Conjugation: New Developments in Enzymatic Processes to Deliver Functionalized Proteins on the Commercial Scale. *Chem. Rev.* (2017). doi:10.1021/acs.chemrev.6b00832
20. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug. Chem.* **24,** 1277–1294 (2013).
21. Ott, W., Jobst, M. A., Schoeler, C., Gaub, H. E. & Nash, M. A. Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox. *J.*

Struct. Biol. **197,** 3–12 (2017).

22. Durner, E., Ott, W., Nash, M. A. & Gaub, H. E. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega* **2,** 3064–3069 (2017).
23. Otten, M. *et al.* From genes to protein mechanics on a chip. *Nat. Methods* **11,** 1127–1130 (2014).
24. Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **5,** 1–8 (2014).
25. Ott, W., Nicolaus, T., Gaub, H. E. & Nash, M. A. Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation. *Biomacromolecules* **17,** 1330−1338 (2016).
26. Ott, W. *et al.* Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano* (2017). doi:10.1021/acsnano.7b02694
27. Catimel, B. *et al.* Kinetic analysis of the interaction between the monoclonal antibody A33 and its colonic epithelial antigen by the use of an optical biosensor. A comparison of immobilisation strategies. *J. Chromatogr. A* **776,** 15–30 (1997).
28. Brogan, K. L., Shin, J. H. & Schoenfisch, M. H. Influence of surfactants and antibody immobilization strategy on reducing nonspecific protein interactions for molecular recognition force microscopy. *Langmuir* **20,** 9729–9735 (2004).
29. Salama-Alber, O. *et al.* Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288,** 16827–16838 (2013).
30. Schoeler, C. *et al.* Mapping Mechanical Force Propagation through Biomolecular Complexes. *Nano Lett.* **15,** 150819094653007 (2015).

31. Verdorfer, T. *et al.* Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.7b07574

32. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–347 (2009).

33. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.* **41,** 207–234 (2005).

Bibliography

[1] S. Pitchiaya and Y. Krishnan, "First blueprint, now bricks: DNA as construction material on the nanoscale", *Chem. Soc. Rev.*, vol. 35, no. 11, pp. 1111–1121, 2006. DOI: 10.1039/B602886C (cit. on p. 1).

[2] L. E. Orgel, "The origin of life – a review of facts and speculations", *Trends in Biochemical Sciences*, vol. 23, no. 12, pp. 491–495, 1998. DOI: 10.1016/S0968-0004(98)01300-0 (cit. on p. 1).

[3] G. van Meer, D. R. Voelker, and G. W. Feigenson, "Membrane lipids: Where they are and how they behave", *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 2, pp. 112–124, 2008. DOI: 10.1038/nrm2330 (cit. on p. 1).

[4] J. Miller, R. Lewontin, W. Gelbart, and A. Griffiths, *Molecular cell biology*. Macmillan Higher Education, 2002, ISBN: 9780716798316 (cit. on p. 1).

[5] J. Howard, *Mechanics of motor proteins and the cytoskeleton*. Sinauer Associates, 2001, ISBN: 9780878933334 (cit. on p. 1).

[6] G. Cooper and R. E. Hausman, *The cell: A molecular approach*. Sinauer Associates, 2015, ISBN: 9781605352909 (cit. on p. 1).

[7] S. Kumar and R. Nussinov, "Close-range electrostatic interactions in proteins", *ChemBioChem*, vol. 3, no. 7, pp. 604–617, 2002. DOI: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X (cit. on p. 1).

[8] T. Hoffmann, K. M. Tych, M. L. Hughes, D. J. Brockwell, and L. Dougan, "Towards design principles for determining the mechanical stability of proteins", *Phys. Chem. Chem. Phys.*, vol. 15, no. 38, pp. 15 767–15 780, 2013. DOI: 10.1039/C3CP52142G (cit. on p. 1).

[9] N. Stephanopoulos and M. B. Francis, "Choosing an effective protein bioconjugation strategy", *Nat. Chem. Biol.*, vol. 7, no. 12, pp. 876–884, 2011. DOI: 10.1038/nchembio.720 (cit. on p. 1).

[10] J. Kalia and R. T. Raines, "Advances in bioconjugation", *Curr. Org. Chem.*, vol. 14, no. 2, pp. 138–147, 2010. DOI: 10.2174/138527210790069839 (cit. on p. 1).

[11] A. Casini, M. Storch, G. S. Baldwin, and T. Ellis, "Bricks and blueprints: Methods and standards for DNA assembly", *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 9, pp. 568–576, 2015. DOI: 10.1038/nrm4014 (cit. on pp. 1, 24).

[12] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, "Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox", *J. Struct. Biol.*, vol. 197, no. 1, pp. 3–12, 2017. DOI: 10.1016/j.jsb.2016.02.011 (cit. on pp. 2, 10, 17, 21, 22, 31).

[13] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, "Ultrastable cellulosome-adhesion complex tightens under load", *Nat. Commun.*, vol. 5, no. 5635, pp. 1–8, 2014. DOI: 10.1038/ncomms6635 (cit. on pp. 2, 9, 21, 45).

[14] C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash, and H. E. Gaub, "Mapping mechanical force propagation through biomolecular complexes", *Nano Lett.*, vol. 15, no. 11, pp. 7370–7376, 2015. DOI: 10.1021/acs.nanolett.5b02727 (cit. on pp. 2, 21, 65).

[15] L. Artzi, E. A. Bayer, and S. Moraïs, "Cellulosomes: Bacterial nanomachines for dismantling plant polysaccharides", *Nat. Rev. Microbiol.*, vol. 15, pp. 83–95, 2017. DOI: 10.1038/nrmicro.2016.164 (cit. on pp. 2, 7, 43).

[16] L. Johanssona, G. Gafvelin, and E. S. Arnéra, "Selenocysteine in proteins - properties and biotechnological use", *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1726, no. 1, pp. 1–13, 2005. DOI: 10.1016/j.bbagen.2005.05.010 (cit. on p. 3).

[17] M. Rother and J. A. Krzycki, "Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea", *Archaea*, vol. 2010, no. 453642, pp. 1–14, 2010. DOI: 10.1155/2010/453642 (cit. on p. 3).

[18] J. Berg, J. Tymoczko, and L. Stryer, *Biochemistry*. W.H. Freeman, 2002, ISBN: 9781464126109 (cit. on p. 3).

[19] F. Sanger, in *Advances in Protein Chemistry*, vol. 7, Academic Press, 1952, pp. 1–67, ISBN: 9780120342884 (cit. on p. 3).

[20] B. A. Shirley, *Protein stability and folding - theory and practice*, ser. Methods in Molecular Biology. Humana Press, 1995, vol. 1, ISBN: 9781592595273 (cit. on p. 3).

[21] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walters, in *Molecular Biology of the Cell*, Garland Publishing, 2002, ISBN: 9780815344643 (cit. on p. 3).

[22] H.-D. Jakubke and H. Jeschkeit, *Aminosäuren, peptide, proteine*. Verlag Chemie, 1982, ISBN: 9783527258925 (cit. on p. 3).

[23] D. Voet, J. G. Voet, and C. W. Pratt, *Fundamentals of biochemistry: Life at the molecular level*. Wiley, 2012, ISBN: 9781118129180 (cit. on p. 3).

[24] G. A. Jeffrey, *An introduction to hydrogen bonding*. Oxford University Press, 1997, ISBN: 9780195095494 (cit. on p. 3).

[25] H. R. Bosshard, D. N. Marti, and I. Jelesarov, "Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings", *J. Mol. Recognit.*, vol. 17, no. 1, pp. 1–16, 2004. DOI: 10.1002/jmr.657 (cit. on p. 4).

[26] J. E. Donald, D. W. Kulp, and W. F. DeGrado, "Salt bridges: Geometrically specific, designable interactions", *Proteins*, vol. 79, no. 3, pp. 898–915, 2011. DOI: 10.1002/prot.22927 (cit. on p. 4).

[27] D. Xu, C.-J. Tsai, and R. Nussinov, "Hydrogen bonds and salt bridges across protein–protein interfaces", *Protein Engineering*, vol. 10, no. 9, pp. 999–1012, 1997. DOI: 10.1093/protein/10.9.999 (cit. on p. 4).

[28] A. D. White, A. J. Keefe, J.-R. Ella-Menye, A. K. Nowinski, Q. Shao, J. Pfaendtner, and S. Jiang, "Free energy of solvated salt bridges: A simulation and experimental study", *J. Phys. Chem. B*, vol. 117, no. 24, pp. 7254–7259, 2013. DOI: 10.1021/jp4024469 (cit. on p. 4).

[29] J. W. Pitera, M. Falta, and W. F. van Gunsteren, "Dielectric properties of proteins from simulation: The effects of solvent, ligands, ph, and temperature", *Biophys. J.*, vol. 80, no. 6, pp. 2546–2555, 2001. DOI: 10.1016/S0006-3495(01)76226-1 (cit. on p. 4).

[30] S. K. Burley and G. A. Petsko, "Aromatic-aromatic interaction: A mechanism of protein structure stabilization", *Science*, vol. 229, no. 4708, pp. 23–28, 1985. DOI: 10.1126/science.3892686 (cit. on p. 4).

[31] E. Lanzarotti, R. R. Biekofsky, D. A. Estrin, M. A. Marti, and A. G. Turjanski, "Aromatic-aromatic interactions in proteins: Beyond the dimer", *J. Chem. Inf. Model.*, vol. 51, no. 7, pp. 1623–1633, 2011. DOI: 10.1021/ci200062e (cit. on p. 4).

[32] S. K. Burley and G. A. Petsko, "Amino-aromatic interactions in proteins", *FEBS Lett.*, vol. 203, no. 2, pp. 139–143, 1986. DOI: 10.1016/0014-5793(86)80730-X (cit. on p. 4).

[33] K. M. Makwana and R. Mahalakshmi, "Implications of aromatic–aromatic interactions: From protein structures to peptide models", *Protein Sci.*, vol. 24, no. 12, pp. 1920–1933, 2015. DOI: 10.1002/pro.2814 (cit. on p. 5).

[34] C. Gomes and P. Wittung-Stafshede, *Protein folding and metal ions: Mechanisms, biology and disease*. CRC Press, 2010, ISBN: 9781439809648 (cit. on p. 5).

[35] S. Nakayama and R. H. Kretsinger, "Evolution of the EF-hand family of proteins", *Annu. Rev. Biophys. Biomol. Struct.*, vol. 23, pp. 473–507, 1994. DOI: 10.1146/annurev.bb.23.060194.002353 (cit. on p. 5).

[36] M. E. Palm-Espling, M. S. Niemiec, and P. Wittung-Stafshede, "Role of metal in folding and stability of copper proteins in vitro", *Biochim. Biophys. Acta*, vol. 1823, no. 9, pp. 1594–1603, 2012. DOI: 10.1016/j.bbamcr.2012.01.013 (cit. on p. 5).

[37] F. H. Arnold and B. L. Haymore, "Engineered metal-binding proteins: Purification to protein folding", *Science*, vol. 252, no. 5014, pp. 1796–1797, 1991. DOI: 10.1126/science.1648261 (cit. on p. 5).

[38] M. M. Harding, "Small revisions to predicted distances around metal sites in proteins", *Acta Crystallographica*, vol. D62, pp. 678–682, 2006. DOI: 10.1107/S0907444906014594 (cit. on p. 5).

[39] W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga, "Disulfide bonds and protein folding", *Biochemistry*, vol. 39, no. 15, pp. 4207–4216, 2000. DOI: 10.1021/bi992922o (cit. on p. 5).

[40] M. Qin, W. Wang, and D. Thirumalai, "Protein folding guides disulfide bond formation", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 36, pp. 11 241–11 246, 2015. DOI: 10.1073/pnas.1503909112 (cit. on p. 5).

[41] C.-H. Wang, C.-C. Huang, L.-L. Lin, and W. Chen, "The effect of disulfide bonds on protein folding, unfolding, and misfolding investigated by FT–Raman spectroscopy", *J. Raman Spectrosc.*, vol. 47, no. 8, pp. 940–947, 2016. DOI: 10.1002/jrs.4935 (cit. on p. 5).

[42] M. T. Neves Petersen, P. H. Jonson, and S. B. Petersen, "Amino acid neighbours and detailed conformational analysis of cysteines in proteins", *Protein Engineering, Design and Selection*, vol. 12, no. 7, pp. 535–548, 1999. DOI: 10.1093/protein/12.7.535 (cit. on p. 5).

[43] D. Chin and A. R. Means, "Calmodulin: A prototypical calcium sensor", *Trends in Cell Biology*, vol. 10, no. 8, pp. 322–328, 2000. DOI: 10.1016/S0962-8924(00)01800-6 (cit. on p. 5).

[44] M. Wilchek, E. A. Bayer, and O. Livnah, "Essentials of biorecognition: The (strept) avidin – biotin system as a model for protein–protein and protein–ligand interaction", *Immunology Letters*, vol. 103, no. 1, pp. 27–32, 2006. DOI: 10.1016/j.imlet.2005.10.022 (cit. on p. 5).

[45] A. N. Schechter, "Hemoglobin research and the origins of molecular medicine", *Blood*, vol. 112, no. 10, pp. 3927–3938, 2008. DOI: 10.1182/blood-2008-04-078188 (cit. on p. 5).

[46] D. A. Fletcher and R. D. Mullins, "Cell mechanics and the cytoskeleton", *Nature*, vol. 463, no. 7280, pp. 485–492, 2010. DOI: 10.1038/nature08908 (cit. on p. 5).

[47] K. A. Eckert and T. A. Kunkel, "DNA polymerase fidelity and the polymerase chain reaction", *Genome Res.*, vol. 1, no. 1, pp. 17–24, 1991. DOI: 10.1101/gr.1.1.17 (cit. on p. 5).

[48] S. J. Remington, "Green fluorescent protein: A perspective", *Protein Sci.*, vol. 20, no. 9, pp. 1509–1519, 2011. DOI: 10.1002/pro.684 (cit. on p. 5).

[49] D. Horton, in *Advances in Carbohydrate Chemistry and Biochemistry*, vol. 64, Academic Press, 2010, ISBN: 9780123808547 (cit. on p. 6).

[50] P. Zugenmaier, *Crystalline cellulose and cellulose derivatives – characterization and structures*. Springer-Verlag, 2008, ISBN: 9783540739340 (cit. on p. 6).

[51] E. Sjostrom, *Wood chemistry fundamentals and applications*. Academic Press, 2013, ISBN: 9780080925899 (cit. on p. 6).

[52] H. P. S. Abdul Khalil, A. H. Bhat, and A. F. Ireana Yusra, "Green composites from sustainable cellulose nanofibrils: A review", *Carbohydr. Polym.*, vol. 87, no. 2, pp. 963–979, 2012. DOI: 10.1016/j.carbpol.2011.08.078 (cit. on p. 6).

[53] H. Wei, Q. Xu, L. E. Taylor 2nd, J. O. Baker, M. P. Tucker, and S.-Y. Ding, "Natural paradigms of plant cell wall degradation", *Curr. Opin. Biotechnol.*, vol. 20, no. 3, pp. 330–338, 2009. DOI: 10.1016/j.copbio.2009.05.008 (cit. on p. 6).

[54] E. A. Bayer, J.-P. Belaich, Y. Shoham, and R. Lamed, "The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides", *Annu. Rev. Microbiol.*, vol. 58, pp. 521–554, 2004. DOI: 10.1146/annurev.micro.57.030502.091022 (cit. on p. 7).

[55] D. E. Koshland Jr., "Stereochemistry and the mechanism of enzymatic reactions", *Biol. Rev. Camb. Philos. Soc.*, vol. 28, no. 4, pp. 416–436, 1953. DOI: 10.1111/j.1469-185X.1953.tb01386.x (cit. on p. 7).

[56] R. Lamed, E. Setter, and E. A. Bayer, "Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum", *J. Bacteriol.*, vol. 156, no. 2, pp. 828–836, 1983 (cit. on p. 7).

[57] K. Cameron, J. Y. Weinstein, O. Zhivin, P. Bule, S. J. Fleishman, V. D. Alves, H. J. Gilbert, L. M. A. Ferreira, C. M. G. A. Fontes, E. A. Bayer, and S. Najmudin, "Combined crystal structure of a type-i cohesin: Mutation and affinity-binding studies reveal structural determinants of cohesin-dockerin specificity", *J. Biol. Chem.*, vol. 290, no. 26, pp. 16 215–16 225, 2015. DOI: 10.1074/jbc.M115.653303 (cit. on p. 7).

[58] J. J. Adams, G. Pal, Z. Jia, and S. P. Smith, "Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 2, pp. 305–310, 2006. DOI: 10.1073/pnas.0507109103 (cit. on p. 7).

[59] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, and E. A. Bayer, "Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses", *J. Biol. Chem.*, vol. 288, no. 23, pp. 16 827–16 838, 2013. DOI: 10.1074/jbc.M113.466672 (cit. on pp. 7, 9).

[60] W. W. Navarre and O. Schneewind, "Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria", *Mol. Microbiol.*, vol. 14, no. 1, pp. 115–121, 1994. DOI: 10.1111/j.1365-2958.1994.tb01271.x (cit. on p. 7).

[61] D. Guillén, S. Sánchez, and R. Rodríguez-Sanoja, "Carbohydrate-binding domains: Multiplicity of biological roles", *Applied Microbiology and Biotechnology*, vol. 85, no. 5, pp. 1241–1249, 2010. DOI: 10.1007/s00253-009-2331-y (cit. on p. 7).

[62] A. B. Boraston, E. Kwan, P. Chiu, R. A. J. Warren, and D. G. Kilburn, "Recognition and hydrolysis of noncrystalline cellulose", *J. Biol. Chem.*, vol. 278, no. 8, pp. 6120–6127, 2003. DOI: 10.1074/jbc.M209554200 (cit. on p. 7).

[63] K. Ruel, Y. Nishiyama, and J.-P. Joseleau, "Crystalline and amorphous cellulose in the secondary walls of arabidopsis", *Plant Sci.*, vol. 193-194, pp. 48–61, 2012. DOI: 10.1016/j.plantsci.2012.05.008 (cit. on p. 7).

[64] S. Jindou, A. Soda, S. Karita, T. Kajino, P. Béguin, J. H. D. Wu, M. Inagaki, T. Kimura, K. Sakka, and K. Ohmiya, "Cohesin-Dockerin interactions within and between clostridium josui and clostridium thermocellum", *J. Biol. Chem.*, vol. 279, no. 11, pp. 9867–9874, 2004. DOI: 10.1074/jbc.M308673200 (cit. on pp. 8, 9).

[65] P. Bule, V. D. Alves, V. Israeli-Ruimy, A. L. Carvalho, L. M. A. Ferreira, S. P. Smith, H. J. Gilbert, S. Najmudin, E. A. Bayer, and C. M. G. A. Fontes, "Assembly of ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes", *Sci. Rep.*, vol. 7, no. 759, pp. 1–14, 2017. DOI: 10.1038/s41598-017-00919-w (cit. on pp. 8, 43).

[66] S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, and H. E. Gaub, "Single-molecule dissection of the high-affinity cohesin-dockerin complex", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 50, pp. 20 431–20 436, 2012. DOI: 10.1073/pnas.1211929109 (cit. on p. 9).

[67] A. Valbuena, J. Oroz, R. Hervás, A. M. Vera, D. Rodríguez, M. Menéndez, J. I. Sulkowska, M. Cieplak, and M. Carrión-Vázquez, "On the remarkable mechanostability of scaffoldins and the mechanical clamp motif", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13 791–13 796, 2009. DOI: 10.1073/pnas.0813093106 (cit. on p. 9).

[68] R. Milo and R. Phillips, *Cell biology by the numbers*. Taylor and Francis Ltd, 2015, ISBN: 9780815345374 (cit. on p. 10).

[69] N. Willet, C. Lamprecht, C. Rankl, M. Rangl, R. Creasey, A. Ebner, N. H. Voelcker, and P. Hinterdorfer, "Molecular manipulation with atomic force microscopy", in, CRC Press, 2012, pp. 3–46, ISBN: 9781439809662 (cit. on pp. 11–15).

[70] J. L. Hutter and J. Bechhoefer, "Calibration of atomic-force microscope tips", *Rev. Sci. Instrum.*, vol. 64, no. 7, pp. 1868–1873, 1993. DOI: 10.1063/1.1143970 (cit. on p. 11).

[71] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, "Entropic elasticity of lambda-phage DNA", *Science*, vol. 265, no. 5178, pp. 1599–1600, 1994. DOI: 10.1126/science.8079175 (cit. on pp. 11, 12).

[72] J. F. Marko and E. D. Siggia, "Stretching DNA", *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, 1995. DOI: 10.1021/ma00130a008 (cit. on pp. 11, 12).

[73] M. Doi and S. F. Edwards, *The theory of polymer dynamics*. oxford university press, 1988, vol. 73, ISBN: 9780198520337 (cit. on p. 12).

[74] M. Rief, J. M. Fernandez, and H. E. Gaub, "Elastically coupled Two-Level systems as a model for biopolymer extensibility", *Phys. Rev. Lett.*, vol. 81, no. 21, pp. 4764–4767, 1998. DOI: 10.1103/PhysRevLett.81.4764 (cit. on p. 12).

[75] B. Zhang and J. S. Evans, "Modeling AFM-induced PEVK extension and the reversible unfolding of Ig/FNIII domains in single and multiple titin molecules", *Biophys. J.*, vol. 80, no. 2, pp. 597–605, 2001. DOI: 10.1016/S0006-3495(01)76040-7 (cit. on p. 12).

[76] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, "Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-initio Theory", *Phys. Rev. Lett.*, vol. 94, no. 4, pp. 048301-1–048301-4, 2005. DOI: 10.1103/PhysRevLett.94.048301 (cit. on p. 12).

[77] L. Livadaru, R. R. Netz, and H. J. Kreuzer, "Stretching Response of Discrete Semiflexible Polymers", *Macromolecules*, vol. 36, no. 10, pp. 3732–3744, 2003. DOI: 10.1021/ma020751g (cit. on p. 13).

[78] W. Baumgartner, P. Hinterdorfer, and H. Schindler, "Data analysis of interaction forces measured with the atomic force microscope", *Ultramicroscopy*, vol. 82, no. 1-4, pp. 85–95, 2000. DOI: 10.1016/S0304-3991(99)00154-0 (cit. on p. 13).

[79] C. Rankl, F. Kienberger, H. Gruber, D. Blaas, and P. Hinterdofer, "Accuracy estimation in force spectroscopy experiments", *Japanese Journal of Applied Physics*, vol. 46, no. 8B, pp. 5536–5539, 2007. DOI: 10.1143/JJAP.46.5536 (cit. on p. 13).

[80] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, "Molecular dynamics study of unbinding of the avidin-biotin complex", *Biophys. J.*, vol. 72, no. 4, pp. 1568–1581, 1997. DOI: 10.1016/S0006-3495(97)78804-0 (cit. on pp. 13, 14).

[81] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds", *Biophys. J.*, vol. 72, no. 4, pp. 1541–1555, 1997. DOI: 10.1016/S0006-3495(97)78802-7 (cit. on pp. 13, 14).

[82] E. Evans and K. Ritchie, "Strength of a weak bond connecting flexible polymer chains", *Biophys. J.*, vol. 76, no. 5, pp. 2439–2447, 1999. DOI: 10.1016/S0006-3495(99)77399-6 (cit. on p. 13).

[83] J. Fritz, A. G. Katopodis, F. Kolbinger, and D. Anselmetti, "Force-mediated kinetics of single p-selectin/ligand complexes observed by atomic force microscopy", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 21, pp. 12 283–12 288, 1998. DOI: 10.1073/pnas.95.21.12283 (cit. on p. 14).

[84] H. Bisswanger, *Enzyme kinetics: Principles and methods*. Wiley, 2009. DOI: 10.1002/9783527622023 (cit. on p. 14).

[85] A. Oberhauser, *Single-molecule studies of proteins*, ser. Biophysics for the Life Sciences. Springer New York, 2012. DOI: 10.1007/978-1-4614-4921-8 (cit. on p. 14).

[86] T. Strunz, K. Oroszlan, I. Schumakovitch, H.-J. Güntherodt, and M. Hegner, "Model energy landscapes and the force-induced dissociation of ligand-receptor bonds", *Biophys. J.*, vol. 79, no. 3, pp. 1206–1212, 2000. DOI: 10.1016/S0006-3495(00)76375-2 (cit. on p. 14).

[87] W. A. Linke and A. Grützner, "Pulling single molecules of titin by AFM–recent advances and physiological implications", *Eur J Physiol*, vol. 456, no. 1, pp. 101–115, 2008. DOI: 10.1007/s00424-007-0389-x (cit. on p. 14).

[88] P. E. Marszalek, H. Li, A. F. Oberhauser, and J. M. Fernandez, "Chair-boat transitions in single polysaccharide molecules observed with force-ramp AFM", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 7, pp. 4278–4283, 2002. DOI: 10.1073/pnas.072435699 (cit. on p. 15).

[89] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, "Stepwise unfolding of titin under force-clamp atomic force microscopy", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 2, pp. 468–472, 2001. DOI: 10.1073/pnas.021321798 (cit. on p. 15).

[90] G. I. Bell, "Models for the specific adhesion of cells to cells", *Science*, vol. 200, no. 4342, pp. 618–627, 1978. DOI: 10.1126/science.347575 (cit. on p. 15).

[91] H. Grubmüller, B. Heymann, and P. Tavan, "Ligand binding: Molecular mechanics calculation of the Streptavidin-Biotin rupture force", *Science*, vol. 271, no. 5251, pp. 997–999, 1996. DOI: 10.1126/science.271.5251.997 (cit. on p. 15).

[92] E. Evans, "Probing the relation between force–lifetime–and chemistry in single molecular bonds", *Annu. Rev. Biophys. Biomol. Struct.*, vol. 30, pp. 105–128, 2001. DOI: 10.1146/annurev.biophys.30.1.105 (cit. on p. 15).

[93] E. M. Puchner, G. Franzen, M. Gautel, and H. E. Gaub, "Comparing proteins by their unfolding pattern", *Biophys. J.*, vol. 95, no. 1, pp. 426–434, 2008. DOI: 10.1529/biophysj.108.129999 (cit. on p. 16).

[94] L. F. Milles, E. A. Bayer, M. A. Nash, and H. E. Gaub, "Mechanical stability of a High-Affinity toxin anchor from the pathogen clostridium perfringens", *J. Phys. Chem. B*, vol. 121, no. 15, pp. 3620–3625, 2017. DOI: 10.1021/acs.jpcb.6b09593 (cit. on p. 17).

[95] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy", *Nat. Methods*, vol. 5, no. 6, pp. 491–505, 2008. DOI: 10.1038/nmeth.1218 (cit. on p. 18).

[96] O. E. Farrance, E. Paci, S. E. Radford, and D. J. Brockwell, "Extraction of accurate biomolecular parameters from single-molecule force spectroscopy experiments", *ACS Nano*, vol. 9, no. 2, pp. 1315–1324, 2015. DOI: 10.1021/nn505135d (cit. on p. 18).

[97] F. Oesterhelt, M. Rief, and H. E. Gaub, "Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water", *New J. Phys.*, vol. 1, no. 6, pp. 1–11, 1999. DOI: 10.1088/1367-2630/1/1/006 (cit. on pp. 18, 177).

[98] Z. Tong, A. Mikheikin, A. Krasnoslobodtsev, Z. Lv, and Y. L. Lyubchenko, "Novel polymer linkers for single molecule AFM force spectroscopy", *Methods*, vol. 60, no. 2, pp. 161–168, 2013. DOI: 10.1016/j.ymeth.2013.02.019 (cit. on p. 19).

[99] W. Ott, T. Nicolaus, H. E. Gaub, and M. A. Nash, "Sequence-Independent Cloning and Post-Translational Modification of Repetitive Protein Polymers through Sortase and Sfp-Mediated Enzymatic Ligation", *Biomacromolecules*, vol. 17, no. 4, pp. 1330–1338, 2016. DOI: 10.1021/acs.biomac.5b01726 (cit. on pp. 19, 179).

[100] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, "Elastin-like polypeptide linkers for Single-Molecule force spectroscopy", *ACS Nano*, vol. 11, no. 6, pp. 6346–6354, 2017. DOI: 10.1021/acsnano.7b02694 (cit. on pp. 19, 215).

[101] J. A. MacKay, D. J. Callahan, K. N. FitzGerald, and A. Chilkoti, "Quantitative model of the phase behavior of recombinant ph-responsive elastin-like polypeptides", *Biomacromolecules*, vol. 11, no. 11, pp. 2873–2879, 2010. DOI: 10.1021/bm100571j (cit. on p. 19).

[102] D. E. Meyer and A. Chilkoti, "Purification of recombinant proteins by fusion with thermally-responsive polypeptides", *Nat. Biotechnol.*, vol. 17, no. 11, pp. 1112–1115, 1999. DOI: 10.1038/15100 (cit. on p. 20).

[103] J. L. Zimmermann, T. Nicolaus, G. Neuert, and K. Blank, "Thiol-based, site-specific and covalent immobilization of biomolecules for single-molecule experiments", *Nat. Protoc.*, vol. 5, no. 6, pp. 975–985, 2010. DOI: 10.1038/nprot.2010.49 (cit. on p. 22).

[104] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, and C. T. Walsh, "Genetically encoded short peptide tag for versatile protein labeling by sfp phosphopantetheinyl transferase", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 44, pp. 15 815–15 820, 2005. DOI: 10.1073/pnas.0507705102 (cit. on p. 23).

[105] B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, and D. R. Liu, "Reprogramming the specificity of sortase enzymes", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 37, pp. 13 343–13 348, 2014. DOI: 10.1073/pnas.1411179111 (cit. on p. 23).

[106] E. Durner, W. Ott, M. A. Nash, and H. E. Gaub, "Post-Translational Sortase-Mediated attachment of High-Strength force spectroscopy handles", *ACS Omega*, vol. 2, no. 6, pp. 3064–3069, 2017. DOI: 10.1021/acsomega.7b00478 (cit. on pp. 23, 245).

[107] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy, and M. Howarth, "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 12, E690–E697, 2012. DOI: 10.1073/pnas.1115485109 (cit. on p. 23).

[108] H. Dietz and M. Rief, "Protein structure by mechanical triangulation", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 5, pp. 1244–1247, 2006. DOI: 10.1073/pnas.0509217103 (cit. on p. 23).

[109] A. J. de Graaf, M. Kooijman, W. E. Hennink, and E. Mastrobattista, "Nonnatural amino acids for site-specific protein conjugation", *Bioconjug. Chem.*, vol. 20, no. 7, pp. 1281–1295, 2009. DOI: 10.1021/bc800294a (cit. on p. 23).

[110] C. Engler, R. Kandzia, and S. Marillonnet, "A one pot, one step, precision cloning method with high throughput capability", *PLoS One*, vol. 3, no. 11, e3647, 2008. DOI: 10.1371/journal.pone.0003647 (cit. on pp. 24, 177).

[111] D. G. Gibson, L. Young, R.-Y. Chuang, C. J. Venter, C. A. Hutchison III, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases", *Nat. Methods*, vol. 6, no. 5, pp. 343–345, 2009. DOI: 10.1038/nmeth.1318 (cit. on pp. 24, 25).

[112] B. Müller-Hill, *The lac operon: A short history of a genetic paradigm*. Walter de Gruyter, 1996, ISBN: 9783110148305 (cit. on p. 26).

[113] F. Studier and B. A. Moffatt, "Use of bacteriophage t7 rna polymerase to direct selective high-level expression of cloned genes", *Journal of Molecular Biology*, vol. 189, no. 1, pp. 113–130, 1986. DOI: 10.1016/0022-2836(86)90385-2 (cit. on p. 26).

[114] F. W. Studier, A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff, "Use of T7 RNA polymerase to direct expression of cloned genes", *Methods Enzymol.*, vol. 185, pp. 60–89, 1990. DOI: 10.1016/0076-6879(90)85008-C (cit. on p. 26).

[115] A. H. Rosenberg, B. N. Lade, C. Dao-shan, S.-W. Lin, J. J. Dunn, and F. Studier, "Vectors for selective expression of cloned dnas by t7 rna polymerase", *Gene*, vol. 56, no. 1, pp. 125–135, 1987. DOI: 10.1016/0378-1119(87)90165-X (cit. on p. 26).

[116] F. W. Studier, "Protein production by auto-induction in high density shaking cultures", *Protein Expression Purif.*, vol. 41, no. 1, pp. 207–234, 2005. DOI: 10.1016/j.pep.2005.01.016 (cit. on p. 27).

[117] W. F. Loomis Jr and B. Magasanik, "Glucose-lactose diauxie in escherichia coli", *J. Bacteriol.*, vol. 93, no. 4, pp. 1397–1401, 1967 (cit. on p. 27).

[118] J. A. Bornhorst and J. J. Falke, "Purification of proteins using polyhistidine affinity tags", in *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification*, ser. Methods in Enzymology, vol. 326, Academic Press, 2000, pp. 245–254, ISBN: 9780080496818 (cit. on p. 27).

[119] T. G. M. Schmidt and A. Skerra, "The strep-tag system for one-step purification and high-affinity detection or capturing of proteins", *Nat. Protoc.*, vol. 2, no. 6, pp. 1528–1535, 2007. DOI: 10.1038/nprot.2007.209 (cit. on p. 27).

[120] A. Zemella, L. Thoring, C. Hoffmeister, and S. Kubick, "Cell-Free protein Synthesis:Pros and cons of prokaryoticand eukaryotic systems", *ChemBioChem*, vol. 16, no. 17, pp. 2420–2431, 2015. DOI: 10.1002/cbic.201500340 (cit. on p. 27).

[121] G. Rosenblum and B. S. Cooperman, "Engine out of the chassis: Cell-free protein synthesis and its uses", *FEBS Lett.*, vol. 588, no. 2, pp. 261–268, 2014. DOI: 10.1016/j.febslet.2013.10.016 (cit. on p. 27).

[122] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda, "Cell-free translation reconstituted with purified components", *Nat. Biotechnol.*, vol. 19, no. 8, pp. 751–755, 2001. DOI: 10.1038/90802 (cit. on p. 27).

[123] Y. Endo and T. Sawasaki, "High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system", *Biotechnol. Adv.*, vol. 21, no. 8, pp. 695–713, 2003. DOI: 10.1016/S0734-9750(03)00105-8 (cit. on p. 27).

[124] F. Katzen, G. Chang, and W. Kudlicki, "The past, present and future of cell-free protein synthesis", *Trends Biotechnol.*, vol. 23, no. 3, pp. 150–156, 2005. DOI: 10.1016/j.tibtech.2005.01.003 (cit. on p. 27).

[125] M. L. Hughes and L. Dougan, "The physics of pulling polyproteins: A review of single molecule force spectroscopy using the AFM to study protein unfolding", *Rep. Prog. Phys.*, vol. 79, no. 7, p. 076 601, 2016. DOI: 10.1088/0034-4885/79/7/076601 (cit. on p. 29).

[126] R. H. Doi and A. Kosugi, "Cellulosomes: Plant-cell-wall-degrading enzyme complexes", *Nat. Rev. Microbiol.*, vol. 2, no. 7, pp. 541–551, 2004. DOI: 10.1038/nrmicro925 (cit. on p. 43).

[127] M. A. Nash, S. P. Smith, C. M. G. A. Fontes, and E. A. Bayer, "Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes", *Curr. Opin. Struct. Biol.*, vol. 40, pp. 89–96, 2016. DOI: 10.1016/j.sbi.2016.08.002 (cit. on p. 43).

[128] A. L. Carvalho, F. M. V. Dias, T. Nagy, J. A. M. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. A. Ferreira, M. J. Romão, C. M. G. a. Fontes, and H. J. Gilbert, "Evidence for a dual binding mode of dockerin modules to cohesins", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 9, pp. 3089–3094, 2007. DOI: 10.1073/pnas.0611173104 (cit. on p. 43).

[129] M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, "Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy", *Elife*, vol. 4, e10319, 2015. DOI: 10.7554/eLife.10319 (cit. on p. 95).

[130] T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub, "Combining in vitro and in silico Single-Molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics", *J. Am. Chem. Soc.*, 2017. DOI: 10.1021/jacs.7b07574 (cit. on p. 115).

[131] I. Hunt, "From gene to protein: A review of new and enabling technologies for multi-parallel protein expression", *Protein Expr. Purif.*, vol. 40, no. 1, pp. 1–22, 2005. DOI: 10.1016/j.pep.2004.10.018 (cit. on p. 153).

[132] S. J. Maerkl and S. R. Quake, "A systems approach to measuring the binding energy landscapes of transcription factors", *Science*, vol. 315, no. 5809, pp. 233–237, 2007. DOI: 10.1126/science.1131007 (cit. on p. 153).

[133] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub, "From genes to protein mechanics on a chip", *Nat. Methods*, vol. 11, no. 11, pp. 1127–1130, 2014. DOI: 10.1038/nmeth.3099 (cit. on p. 155).

[134] T. Kowalczyk, K. Hnatuszko-Konka, A. Gerszberg, and A. K. Kononowicz, "Elastin-like polypeptides as a promising family of genetically-engineered protein based polymers", *World Journal of Microbiology and Biotechnology*, vol. 30, no. 8, pp. 2141–2152, 2014. DOI: 10.1007/s11274-014-1649-5 (cit. on p. 177).

List of Figures

Acknowledgements

During my research at the Gaub Lab I received a lot of support, and for that I am very grateful. I thank all lab members for contributing to such a friendly and fruitful working environment.
Thank you **Hermann Gaub**, for opening the door and welcoming me in the world of physics. I am very grateful for your guidance of my work in your group, the vivid scientific discussions, your ongoing support on experiments and your input for future Postdoc-Positions.
I thank **Diana Pippig** for helping me through the transition from biochemistry into biophysics and your ongoing help in biochemistry.
Thank you, **Edward Bayer**, for introducing me to the cellulosome, natures most intriguing model system to study protein network dynamics.
The Nashcats: **Michael Nash**, I thank you for forming the subgroup, fostering all our ideas from seed to harvest and your ongoing enthusiasm in our projects from Basel.
Markus Jobst, **Tobias Verdorfer**, **Lukas Milles** and **Ellis Durner** - I thank you for being my guides on the adventure physics, providing me with all tools necessary and also some that are not that useful. The atmosphere in the office is certainly unique and a most memorable buffer against (scientific) fallbacks, a source of inspiration and a fundament for future friendship. I thank you for teaching me AFM theory, introducing me to AFM operations, for providing the code for AFM measurements *via* the worlds presumably tiniest snail, for the automated data analysis replacing the monkey button and overhauling and automation of the AFMs. I enjoyed the scientific discussions and thank you for your patience during proof-reading of the thesis.
I thank **Constantin Schöler** for helpful and encouraging discussions about the ongoing thesis writing.
I thank **Thomas Nicolaus** for your patience with ELP cloning, not to mention about the hundreds of other clones you created, boosting and enabling my work. Your omnipresent good mood is infectious to everyone entering the lab, which might be conserved by your music.
Angelika Kardinal, I thank you for your work on perceived thousands of plasmid isolations forming the basis of my research. Your good mood and encouraging words throughout our lab bench neighborhood provided an ideal working atmosphere.
I thank **Fabian Baumann** for his input in all fluorescence- and TIRF-based questions as well as his support and good spirits throughout the thesis.
I thank **Magnus Bauer** for his support in the lab and the discussions about science, life and future.
I am thanking **Willem Vanderlinden** for imaging the unstructured ELPs, his interest, input and support in ongoing science projects and the provision of Belgian beers.
Linda Brützel, I thank you for characterizing ELPs with SAXS and your help with all questions around thesis organization.
I thank **Sylvia Kreuzer**, for organizing everything, your tremendous help on every administrative matter and of course providing sweets to brighten every visit in your office.

I thank my parents, **Maria** and **Hermann** and my brothers **Andreas**, **Michael** and **Josef**. Without your support and motivation I would not be where I am now.
Beatrice I am deeply grateful, not only your moral, scientific and proof-reading support allowed this thesis to evolve. Thank you for listening to my (science-related) concerns, brightening my mood when something failed and escaping with me from the scientific world whenever necessary. Thank you for having my back.
Finally, I would like to thank **Steve Gunn**, **Sunil Patel** and **Johannes Böttcher** for providing the basic latex-template for this work, under Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0).

Eidesstattliche Erklärung

Ich, Wolfgang Bernhard Ott, versichere die Arbeit

"Single Molecule Force Spectroscopy
with Biological Tools"

selbstständig angefertigt zu haben und nur die dazu im Literaturverzeichnis angegebenen Quellen benutzt zu haben.

München, 15.11.2017

_____